



RATAN TATA
LIBRARY

STATISTICAL METHODS

STATISTICAL METHODS

APPLIED TO EXPERIMENTS IN
AGRICULTURE AND BIOLOGY

By GEORGE W. SNEDECOR

DIRECTOR OF THE STATISTICAL LABORATORY OF
IOWA STATE COLLEGE AND HEAD OF THE STATISTICAL SECTION OF
THE IOWA AGRICULTURAL EXPERIMENT STATION

FOURTH EDITION



THE IOWA STATE COLLEGE PRESS

AMES, IOWA

Copyright, 1937, 1938, 1940, and 1946, by
THE IOWA STATE COLLEGE PRESS

Second Edition, 1938

Third Edition, 1940

Fourth Edition, 1946

Second Printing, 1946

Third Printing, 1948

PRINTED AT THE IOWA STATE COLLEGE PRESS
AMES, IOWA, U.S.A.

TO
GERTRUDE

PREFACE

The beginner in experimentation too often finds himself supplied with a pair of elaborate mechanisms. In the one hand is a mass of data demanding simplification and interpretation, while in the other is a complex statistical methodology said to be necessary to research. How shall the two be geared together? Since the data can be only inefficiently utilized without statistical method, and since method is futile until applied to data, it seems strange that greater effort has not been made to unite the two. For those of some experience there are adequate texts and journal articles. It is the novice to whose needs this book is directed. It is hoped that he may be furnished with a smoothly working combination of experimental data and statistical method.

Like all other sciences, statistics is in a stage of rapid evolution. During the last 20 years new discoveries have swiftly succeeded each other, fruitful syntheses have been effected, novel modes of thought have developed and a whole series of brand new statistical methods have been marketed. The biologist who has not been able to keep abreast of the progress of statistics finds himself a bit confused by the new ideas and technical terms. It is thought that he will welcome a statement of them in a form that will not require too much distraction of his attention from necessary professional duties.

It is a fundamental belief of the author that statistical method can be used competently by scientists not especially trained in mathematics. The conditions surrounding the mathematical theorems can be set forth in terms quite readily understood by the lay reader. Since mastery of two sciences is possible for only few, it is necessary for most of us to advance by cooperation. To the mathematical statistician must be delegated the tasks of developing the theory and devising the methods, accompanying these latter by adequate statements of the limitations on their use. None but the biologist can decide whether the conditions are fulfilled in his experiments and interpret the results. The only mathematics used in this book is arithmetic, supplemented by enough symbolism to make the exposition intelligible.

In the course of the development of each bit of scientific knowledge there comes a time when the experimental techniques must be questioned. Are they adequate to furnish the demanded precision of results? In what respects need they be improved? Is the most hopeful point of attack in the laboratory methods or in the experimental material? Fortunately, statistical methods supply answers, in many cases with little or no extra labor in collecting data, provided only that slight but necessary modifica-

tions be included in the plan of the experiment. Some of these tests of technique can be discussed even in this elementary presentation.

Small sample methods are prerequisite in most biological data. For that reason, they are introduced at the start. The classical theory of large samples receives scant attention. In most places where it is mentioned at all it is introduced as a simplified special case of the small sample.

The arrangement of the material in this text is not so much logical as developmental. The easiest ideas are put first, and only one new concept is presented at a time. The experienced reader will often feel a sense of inadequacy. It is believed that this will disappear as he continues, and that the inexperienced will be inducted with a minimum of difficulty. Numerous examples form an indispensable part of the presentation. In most of them the statistical method, with its meaning, is emphasized, the necessary drudgery of calculation being reduced to the lowest level.

Certain diligent but misguided enthusiasts have brought down upon statistics the opprobrious description, "dry as dust." Of course, one must take into consideration the point of view. Data on golf scores, operations, and babies are arid indeed to the listener, but of absorbing interest to the narrator. We have endeavored to present the subject in a different aspect. Fundamentally, statistics is a mode of thought. Biometrics is a delineation of living things. While the mechanism of description is always likely to be tedious, the effort has been made to emphasize the subject portrayed rather than the technique of the portrayal.

Statistics at Iowa State College is a cooperative enterprise. In a sense the author is merely reporting the thinking of his colleagues. Their interest, advice and help have made possible the experience upon which this book is founded. Their generous contributions of experimental data and technical knowledge will, if I have succeeded in interpreting them adequately, be helpful to others engaged in research.

It is a pleasure to acknowledge the leadership of Professor R. A. Fisher. Even he who runs may read my appreciation of his unifying contributions to statistics. By his residences as guest professor in mathematics at Iowa State College, as well as through his writings, he has exercised a profound influence on the experimental and statistical techniques of the institution. He and his publishers, Messrs. Oliver and Boyd of Edinburgh, have been liberal in permitting the use of tables of functions.

My collaborators in the Statistical Laboratory have been unsparing of their help. To A. E. Brandt, Gertrude M. Cox, H. W. Norton and Mary L. Greenwood, I am indebted for valuable criticisms, suggestions, and computational assistance.

GEORGE W. SNEDECOR

PREFACE TO THE THIRD EDITION

I seize this opportunity of expressing my grateful appreciation of the friendly and helpful comments received from readers in all parts of the world. In this third printing I have corrected a number of errors that have been called to my attention. New material has been added at the end of chapters 6 and 16, while an additional chapter 17 has been written to incorporate methods useful in some of the broader fields of sampling.

My colleague, Professor W. G. Cochran, has given me indispensable help and criticism in revising and extending the text.

G. W. S.

September, 1940

PREFACE TO THE FOURTH EDITION

In this edition the text has been largely rewritten, and the scope has been widened as follows: (i) greater emphasis has been placed on the theoretical conditions in which the various statistical methods have validity, and concurrently (ii) on the conduct of the experiment so as to incorporate in the data the information desired; (iii) estimates and fiducial statements have been brought into equal prominence with tests of hypotheses; (iv) there is increased reliance on experimental samplings to exemplify distribution theory; (v) the treatment of correlation and of experimental designs has been expanded; and (vi) the methods for disproportionate subclass numbers have been extended to include all those necessary for ordinary needs.

As before, I have leaned heavily on my colleagues in the Statistical Laboratory and on other members of the Iowa State College staff. My indebtedness to them is gratefully acknowledged.

G. W. S.

January, 1946

A SHORT COURSE IN THE ELEMENTS OF STATISTICAL METHOD

CHAPTER	PAGES
1. <i>Attributes</i>	1-24
2. <i>Individual comparisons</i>	31-46
3. <i>Sampling</i>	54-68
4. <i>Group comparisons</i>	75-82
5. <i>Short-cuts</i>	89-94
6. <i>Regression</i>	103-121
7. <i>Correlation</i>	138-150
10. <i>Analysis of variance</i>	
<i>Single criterion of classification</i>	214-227
11. <i>Analysis of variance</i>	
<i>Two criteria of classification</i>	253-264

TABLE OF CONTENTS

INTRODUCTION	xvi
1. SAMPLING OF ATTRIBUTES	1
A sampling described. Inferences about the population. Definitions. An exercise in sampling. Hypotheses about the population. Chi-square. An experiment in sampling chi-square. Frequency distribution. Graphical representation. Significance. Sample size. Ratios, rates and percentages.	
2. MEASUREMENT DATA. COMPARISON OF INDIVIDUALS	31
The mean as an unbiased estimate. Range, array, standard deviation. Fiducial inference. Test of null hypothesis. Coefficient of variation. Probable deviation. Median, quartiles.	
3. SAMPLING FROM A NORMALLY DISTRIBUTED POPULATION	54
Random samples. Distribution of mean, of variance, of standard deviation and of t . Standard error. Fiducial statement. Test of null hypothesis. Reducing the experimental error.	
4. THE COMPARISON OF TWO GROUPS	75
Groups with equal numbers of individuals. Difficulties met in experiments with groups. Size of sample. Groups with different numbers of individuals. Validity of tests of null hypotheses. Individual vs. group comparison. Statistics and the experiment.	
5. SHORT CUTS AND APPROXIMATIONS	89
Standard deviation. Method for calculating machines. Code numbers. Significant digits. Range/sigma. Common sense in computations. Rates and percentages.	
6. LINEAR REGRESSION	103
Regression containing the origin, homogeneous variance. The regression, $\hat{Y} = a + bX$. Comparison of individuals and of means. Standard deviation from regression. Fiducial statements and tests of null hypotheses. Short cut methods. Statistical control. Partitioning the sum of squares of the dependent variate.	
7. CORRELATION	138
Correlation coefficient. Calculation with code numbers. Relation between coefficients of correlation and regression. Sampling variation of correlation coefficient. Common elements. Estimation, fiducial statements and tests of null hypotheses. Graph of z . Correlation and regression. Correlation, common elements and regression. Correlation and differences. Correlations with sums and ratios—spurious correlation. Correlation due to common causes—nonsense correlations. Correlation of ranks.	

8. LARGE SAMPLE METHODS	169
Computation of mean and standard deviation. Test of symmetry. Test of normality. Graph of the normal curve. Table of normal frequency distribution. Calculation of correlation coefficient in two-way table. Large sample approximations.	
9. ENUMERATION DATA WITH MORE THAN ONE DEGREE OF FREEDOM	188
An experiment with more than one sample. An experiment with more than two classes of individuals. Adjustment of chi-square. Testing chi-square with more than 30 degrees of freedom. Test of independence. The $2 \times 2 \times 2$ table. The $R \times C$ table. Special methods in $R \times 2$ table. Test of experimental technique. Sampling from the Poisson distribution.	
10. ANALYSIS OF VARIANCE IN TWO OR MORE GROUPS OF MEASUREMENT DATA	214
Multiple samples from a common population. Calculation of analysis of variance. Test of null hypothesis—the variance ratio, F . Analysis of variance in two subsamples. Estimates of variance. Subsamples with different numbers of individuals. Analysis of variance of subsample means. Intraclass correlation. Efficiency of an experiment. Tests of homogeneity of variance.	
11. ANALYSIS OF VARIANCE WITH TWO OR MORE CRITERIA OF CLASSIFICATION	253
Use of knowledge about outcome. Randomized blocks. The nature of discrepancy. Estimates of variance. Tests of null hypotheses. Missing data. Latin square. More than one item in the subclasses. Proportional subclass numbers. Disproportionate subclass numbers. Experiments in several places or years. Three criteria of classification. Split plot experiment. Perennial plants in randomized blocks. Correlation between mean and variance.	
12. COVARIANCE	318
Covariance in groups. Test of significance of differences among adjusted means. Covariance in randomized blocks; in Latin square.	
13. MULTIPLE REGRESSION AND COVARIANCE	340
Three variates in a single group; the correlation method. The multiple regression equation. Deviations from regression. The multiple correlation coefficient, R . Multiple covariance in groups; in randomized blocks. Partial correlation. Four or more variates. The direct method of calculation. Gauss multipliers; tests of significance and fiducial limits. Deletion of a variate. Relation between the correlation and direct methods.	
14. CURVILINEAR REGRESSION	374
Exponential growth curve. Second degree polynomial. Test of deviation from linear regression. Fitting orthogonal polynomials.	
15. INDIVIDUAL DEGREES OF FREEDOM	400
Short method of calculating mean square. Orthogonal sets of comparisons. Nonorthogonal sets. Comparisons not planned in the experimental design. Individual comparisons in regression. Applications to randomized blocks. Switchback or reversal experiments. Factorial experiments.	

16. BINOMIAL AND POISSON DISTRIBUTIONS	431
Binomial distribution. Comparison of sample distribution with binomial. Test of homogeneity. Poisson distribution. Comparison of means in samples from Poisson distributions. Transformations of data for the <i>F</i> test.	
17. DESIGN AND ANALYSIS OF SAMPLINGS	453
Populations. Sampling from a homogeneous population. Size of sample. Sampling from small populations. Sampling from two or more homoge- neous populations. Analysis of the data. The structure of sampling in- vestigations. Effectiveness of stratification.	
INDEX OF MATHEMATICAL TABLES	477
GENERAL INDEX AND INDEX OF SYMBOLS	477

INTRODUCTION

There is a fascination in experimental work felt even by those who are not so fortunate as to follow it as a vocation. In searching the unknown for new truths there is mystery, there is adventure, there is the thrill of discovery. As most of us know, also, there are grinding routine and disappointment. Research is a hard mistress, but her rewards bring satisfactions that few of us would forego.

The experiment is the distinctive tool of science. As old as man himself, its earliest forms were no more than observations of the results following an innovation, usually accidental. Attracted by a lush growth of grain, the primitive scientist remembered burying a fish at the spot. So may have begun the fertilization of crops. The story goes that, to frighten away thieves from the grapes in Medoc, a mixture of lime and copper sulfate was sprayed on the posts supporting the vines. A keen observer, Millardet, noticed that the nearby leaves were free from mildew and suggested spraying the vines with "Bordeaux Mixture." These crude experiments were the prototypes of some of our modern demonstrations. You are shown two animals in a cage. One is sleek, alert and active, while the other is scrawny and nearly blind, the result, you are told, of deficiency of vitamin A in the ration. Again, you visit a farm where two adjacent plots of acid soil were sown with clover. One plot, treated with lime and inoculated with nodule bacteria, has a uniform stand of vigorously growing clover while the other is occupied chiefly by weeds. When differences are so great, no more elaborate experiment is needed to prove the benefit of the treatment.

Scarcely are the more obvious questions answered, however, when the less apparent ones begin to clamor for investigation. Which foods contain the best supplies of vitamins, and what rate of application of lime will yield the maximum returns? The solution of these problems requires experiments which, while becoming more complex, may be so planned as to become concurrently more efficient.

The results of experiments are usually recorded in number symbols. The reduction of these data to summary forms, with the accompanying tests for reliability, gives rise to a branch of arithmetic called *statistical method*. Inevitably, the necessity for statistics and the intricacy of the methods are augmented by every advance toward more critical experimentation. An adequate experimental design involves not only a satisfactory plan for conducting the trial, but also an appropriate statistical

method for evaluating the results. The two can ill be divorced. United, they lead to conclusions that lend themselves to statement in categorical terms.

To be prepared for research in biology, one should master the simpler techniques of experimental design. Not only must the many precautions be observed to insure the success of the laboratory or field work, but also the appropriate statistical method must be available for reducing and testing the data. The more elaborate designs may not be needed at first, or indeed ever. However, they are all at the command of the biologist who has a thorough knowledge of the fundamentals of statistical method.

Chapter 1

SAMPLING OF ATTRIBUTES

1.1—The problems. Sampling of public opinion enjoys at the moment a popularity and confidence that are well deserved. Less in the public eye but no less useful are numerous other practical samplings—devices for learning about large masses by observing a few individuals. A dozen rats, for example, may suffice to disclose useful facts concerning a population that runs into millions. A carload of coal is accepted or rejected on the evidence gained from testing only a few pounds. The physician makes inferences about a patient's blood through examination of a single drop. This practice of sampling, so widespread and so fraught with consequences, deserves more than passing scrutiny.

If correct conclusions are to be drawn, the sample must be representative of the population. Keen thinking and long experience have yielded extensive knowledge of how samples should be gathered for maximum information. But despite the best efforts of the investigator, he always faces the possibility that the results from any one sample may be far from the true values in the aggregate. In sections to follow, you will be confronted by the rather disconcerting variability of samples drawn from a common source. Taking precautions to make them representative is, indeed, one part of the sampler's obligation; but he must learn, also, as he makes his inferences to allow for ever-present sampling fluctuation. Here then are the twin problems that face the statistician: the collection of appropriate samples and the drawing of valid conclusions from them.

Sampling may be far-flung, as in the effort to learn how an election is going; or it may be confined to a greenhouse bench. While most of this book is devoted to this latter, the experimental type, it is easier to start by giving some account of the former.

1.2—A sampling described. To focus attention, let us study a project undertaken in 1942, the objective being to estimate the effect of labor shortage on the production of pigs in eastern South Dakota during the spring of 1943. The sampling was done in this manner: using two randomly selected assessors' books from each of 38 counties, we mailed a questionnaire to every fifth farmer in these lists. A sample of 1,000 replies was chosen so as to give uniform representation to all sizes of farms. Of these 1,000 farmers, 43% indicated that extra help would not enable them to increase their farrowing of spring pigs.

Let us examine this procedure in some detail. As you read on, you will find insistent emphasis on the necessity of randomness in sampling to

simplify the reaching of correct inferences about the population sampled, the population in this case being the 54,000 farmers in eastern South Dakota. What is a random sample? The dictionary defines it as one in which the selection of the individuals is left to chance. Some mechanical method is desirable, such as the drawing of numbered capsules from a container in which they are thoroughly mixed. Every capsule should have the same likelihood of being drawn; that is, there should be no favoritism or *bias*.

From this description, you will see that our sample of farms is not simply random. In the first place, the assessors' books were not taken entirely by chance, but two were drawn at random from each county. Such is called *stratified random sampling*, and will be discussed more at length in subsequent chapters. In the second place, we used a systematic scheme of taking the names from the books, every fifth being selected. This, though often convenient, may lead to non-random sampling if there is any order to the way the names are entered in the list. That is, if the entry is by community, or alphabetical, the systematic selection may not yield a random sample. On the other hand, if the first name is chosen by lot and every fifth taken thereafter, then every farm has the same chance of appearing in the sample; and if there is no grouping of the names in the books by family, by community, or otherwise, the resulting sample should be random. Finally, the sample of 1,000 now under discussion was systematically spread over all farm sizes because it was thought that size might be related to labor needs. This is another kind of stratification for which appropriate methods will be developed in chapter 6. You see, we have used some devices which do not violate the requirement of randomness and which increase the precision of the sampling, but you will have to wait to learn how this greater precision can be utilized. For the present we shall assume that we have a wholly random sample.

1.3—Inferences about the population. As noted before, 430 or 43% of the sample of 1,000 farmers reported that they could not increase the spring farrowing of pigs with additional help. What inferences can be made about the sampled population of 54,000 farmers? The first one is this: of the fraction of farmers in eastern South Dakota who could not increase their pig crops even if extra labor were available, 43% is an unbiased estimate. In other words, the sample percentage, 43%, is an unbiased estimate of the population percentage. This inference is usually taken for granted. Most people pass from sample fact to inference about the population with scarcely a thought. Logically, the two concepts are distinct, and unless the sampling is random the inference may not be true.

The second inference about the population is made by use of table 1.1(2)*. In the first part of the table, indicated by 95% in the heading,

* Reference 2 at the end of the chapter.

look across the top line to the sample size of 1,000, then down the column, X/n , to 43%. At the intersection of the line and column you will see the figures 40 and 46. The second inference is that between 40% and 46% of the population of 54,000 farmers probably cannot utilize extra help to increase the spring farrowings. Thus we have set bounds to the inaccuracy of the sample estimate. First we said that the observed percentage is an unbiased estimate of the population value but gave no indication of the amount by which it might be in error. Now we assert the likelihood that the true percentage is not more than 3 ~~either~~ way from 43%.

Since these concepts lie at the foundation of sampling theory, let us illustrate them in another fashion. Imagine a bin filled with beans, some white and some colored, thoroughly mixed. Dip out a scoopful of them at random, count the number of each color and calculate the percentage of white, say 24%. This is an estimate of the fraction of white beans in the bin. But how close an estimate is it? That's where the second inference comes in. If there were 250 beans in the scoop, we look at the table and say that the portion of white beans in the bin is probably between 19% and 30% of the total. What is the nature of the probability involved in this statement? It is this: in this scoopful we may have happened upon an unusual concentration of white beans, or of colored, despite our efforts at complete mixing and random sampling. The table heading indicates the chance of a correct inference as 95 in 100 or, as it is often written, 19:1. This means that if the scooping is repeated many times, each time with a new estimate and a correspondingly new *confidence interval*, and if each time we use the table to make an appropriate confidence statement about the population percentage, we may expect to be right in about 19 of every 20 trials.

Summarizing the farm labor sampling, we have estimated the proportion of farmers in eastern South Dakota not able to utilize extra help as 43%, and we have said that the true percentage lies between 40% and 46% unless a one-in-twenty chance has occurred in the sampling. If you feel unsafe in making inferences with a 5% probability of being wrong, you may use the second part of the table in which $P = 0.99$, setting the wider interval, 39% to 47%. If one says that the population percentage lies within these limits, he will be right unless a one-in-a-hundred chance has come off.

You now realize why it is necessary to distinguish sharply the sample estimate, known exactly, from the population ratio, usually unknown but confidently asserted to lie within the tabled interval. The confusion of these two concepts lies at the bottom of a lot of muddy thinking about statistics.

Another form of the inferences can be used with samples such as ours, drawn from a population whose size is known, 54,000 farms. Not only the estimate, 43%, but both of the confidence limits, 40% and 46%, may be *expanded* by multiplication, 43% of 54,000 = 23,200, etc. (Avoid the

TABLE 1.1
95% CONFIDENCE INTERVAL (PERCENT) FOR BINOMIAL DISTRIBUTION

Number Observed X	Size of Sample						Fraction Observed X/n	Size of Sample	
	10	15	20	30	50	100		250	1000
0	0 31	0 22	0 17	0 12	0 07	0 4	.00	0 1	0 0
1	0 45	0 32	0 25	0 17	0 11	0 5	.01	0 4	0 2
2	3 56	2 40	1 31	1 22	0 14	0 7	.02	1 5	1 3
3	7 65	4 48	3 38	2 27	1 17	1 8	.03	1 6	2 4
4	12 74	8 55	6 44	4 31	2 19	1 10	.04	2 7	3 5
5	19 81	12 62	9 49	6 35	3 22	2 11	.05	3 9	4 7
6	26 88	16 68	12 54	8 39	5 24	2 12	.06	3 10	5 8
7	35 93	21 73	15 59	10 43	6 27	3 14	.07	4 11	6 9
8	44 97	27 79	19 64	12 46	7 29	4 15	.08	5 12	6 10
9	55 100	32 84	23 68	15 50	9 31	4 16	.09	6 13	7 11
10	69 100	38 88	27 73	17 53	10 34	5 18	.10	7 14	8 12
11		45 92	32 77	20 56	12 36	5 19	.11	7 16	9 13
12		52 96	36 81	23 60	13 38	6 20	.12	8 17	10 14
13		60 98	41 85	25 63	15 41	7 21	.13	9 18	11 15
14		68 100	46 88	28 66	16 43	8 22	.14	10 19	12 16
15		78 100	51 91	31 69	18 44	9 24	.15	10 20	13 17
16			56 94	34 72	20 46	9 25	.16	11 21	14 18
17			62 97	37 75	21 48	10 26	.17	12 22	15 19
18			69 99	40 77	23 50	11 27	.18	13 23	16 21
19			75 100	44 80	25 53	12 28	.19	14 24	17 22
20			83 100	47 83	27 55	13 29	.20	15 26	18 23
21				50 85	28 57	14 30	.21	16 27	19 24
22				54 88	30 59	14 31	.22	17 28	19 25
23				57 90	32 61	15 32	.23	18 29	20 26
24				61 92	34 63	16 33	.24	19 30	21 27
25				65 94	36 64	17 35	.25	20 31	22 28
26				69 96	37 66	18 36	.26	20 32	23 29
27				73 98	39 68	19 37	.27	21 33	24 30
28				78 99	41 70	19 38	.28	22 34	25 31
29				83 100	43 72	20 39	.29	23 35	26 32
30				88 100	45 73	21 40	.30	24 36	27 33
31					47 75	22 41	.31	25 37	28 34
32					50 77	23 42	.32	26 38	29 35
33					52 79	24 43	.33	27 39	30 36
34					54 80	25 44	.34	28 40	31 37
35					56 82	26 45	.35	29 41	32 38
36					57 84	27 46	.36	30 42	33 39
37					59 85	28 47	.37	31 43	34 40
38					62 87	28 48	.38	32 44	35 41
39					64 88	29 49	.39	33 45	36 42
40					66 90	30 50	.40	34 46	37 43
41					69 91	31 51	.41	35 47	38 44
42					71 93	32 52	.42	36 48	39 45
43					73 94	33 53	.43	37 49	40 46
44					76 95	34 54	.44	38 50	41 47
45					78 97	35 55	.45	39 51	42 48
46					81 98	36 56	.46	40 52	43 49
47					83 99	37 57	.47	41 53	44 50
48					86 100	38 58	.48	42 54	45 51
49					89 100	39 59	.49	43 55	46 52
50					93 100	40 60	.50	44 56	47 53

* If X exceeds 50, read $100 - X$ = number observed and subtract each confidence limit from 100.

† If X/n exceeds 0.50, read $1.00 - X/n$ = fraction observed and subtract each confidence limit from 100.

TABLE 1.1(continued)
99% CONFIDENCE INTERVAL (PER CENT) FOR BINOMIAL DISTRIBUTION

Number Observed X	Size of Sample						Fraction Observed X/n	Size of Sample	
	10	15	20	30	50	100		250	1000
0	0 41	0 30	0 23	0 16	0 10	0 5	.00	0 2	0 1
1	0 54	0 40	0 32	0 22	0 14	0 7	.01	0 5	0 2
2	1 65	1 49	1 39	0 28	0 17	0 9	.02	1 6	1 3
3	4 74	2 56	2 45	1 32	1 20	0 10	.03	1 7	2 4
4	8 81	5 63	4 51	3 36	1 23	1 12	.04	2 9	3 6
5	13 87	8 69	6 56	4 40	2 26	1 13	.05	2 10	3 7
6	19 92	12 74	8 61	6 44	3 29	2 14	.06	3 11	4 8
7	26 96	16 79	11 66	8 48	4 31	2 16	.07	3 13	5 9
8	35 99	21 84	15 70	10 52	6 33	3 17	.08	4 14	6 10
9	46 100	26 88	18 74	12 55	7 36	3 18	.09	5 15	7 12
10	59 100	31 92	22 78	14 58	8 38	4 19	.10	6 16	8 13
11		37 95	26 82	16 62	10 40	4 20	.11	6 17	9 14
12		44 98	30 85	18 65	11 43	5 21	.12	7 18	9 15
13		51 99	34 89	21 68	12 45	6 23	.13	8 19	10 16
14		60 100	39 92	24 71	14 47	6 24	.14	9 20	11 17
15		70 100	44 94	26 74	15 49	7 26	.15	9 22	12 18
16			49 96	29 76	17 51	8 27	.16	10 23	13 19
17			55 98	32 79	18 53	9 29	.17	11 24	14 20
18			61 99	35 82	20 55	9 30	.18	12 25	15 21
19			68 100	38 84	21 57	10 31	.19	13 26	16 22
20			77 100	42 86	23 59	11 32	.20	14 27	17 23
21				45 88	24 61	12 33	.21	15 28	18 24
22				48 90	26 63	12 34	.22	16 30	19 26
23				52 92	28 65	13 35	.23	17 31	20 27
24				56 94	29 67	14 36	.24	18 32	21 28
25				60 96	31 69	15 38	.25	18 33	22 29
26				64 97	33 71	16 39	.26	19 34	22 30
27				68 99	35 72	16 40	.27	20 35	23 31
28				72 100	37 74	17 41	.28	21 36	24 32
29				78 100	39 76	18 42	.29	22 37	25 33
30				84 100	41 77	19 43	.30	23 38	26 34
31					43 79	20 44	.31	24 39	27 35
32					45 80	21 45	.32	25 40	28 36
33					47 82	21 46	.33	26 41	29 37
34					49 83	22 47	.34	26 42	30 38
35					51 85	23 48	.35	27 43	31 39
36					53 86	24 49	.36	28 44	32 40
37					55 88	25 50	.37	29 45	33 41
38					57 89	26 51	.38	30 46	34 42
39					60 90	27 52	.39	31 47	35 43
40					62 92	28 53	.40	32 48	36 44
41					64 93	29 54	.41	33 50	37 45
42					67 94	29 55	.42	34 51	38 46
43					69 96	30 56	.43	35 52	39 47
44					71 97	31 57	.44	36 53	40 48
45					74 98	32 58	.45	37 54	41 49
46					77 99	33 59	.46	38 55	42 50
47					80 99	34 60	.47	39 55	43 51
48					83 100	35 61	.48	40 56	44 52
49					86 100	36 62	.49	41 57	45 53
50					90 100	37 63	.50	42 58	46 54

* If X exceeds 50, read $100 - X$ = number observed and subtract each confidence limit from 100.

† If X/n exceeds 0.50, read $1.00 - X/n$ = fraction observed and subtract each confidence limit from 100.

fallacy of carrying more significant figures than are justified by the factors multiplied. See section 5.6.) We may infer (i) that 23,200 is an unbiased estimate of the number of farmers in eastern South Dakota who cannot use extra labor to increase their pig enterprises, and (ii) that the real number lies between 21,600 and 24,800 unless a one-in-twenty chance has occurred in the sampling. It may be noted that corresponding inferences may be made about the alternative class, those farmers who *could* increase the pig crop if extra labor were available, the estimate being 57% (expanded to 30,800 farmers) with confidence limits, 54% and 60%.

In this brief prevue I have displayed a goodly portion of the wares that the statistician has to offer; the sampling of aggregates, examination of the facts discovered in the sample, and inferences based upon these facts. The remainder of the book is occupied chiefly with more and more intricate applications of these basic processes. Before going further, you may clarify your thinking by working out a few examples.

You will find that examples form an essential part of my presentation of statistics. In each list they are graded so that you may start with the easier. It is suggested that a few in each group be worked after the first reading of the text, reserving the more difficult until experience is enlarged. It is my firmly grounded belief that statistics cannot be mastered without this or similar practice.

EXAMPLE 1.1—Of a sample of 100 seeds drawn at random from a bag of uniformly mixed bluegrass seeds, 92 germinated in a standard test. The seed analyst wishes to make a statement about the proportion of seeds in the bag that will germinate under similar conditions, and he is willing to run the risk of only one per hundred that his sample has misled him. What may he say? Ans. Between 83% and 97% will germinate. Hint: look in the table for $100-92 = 8\%$.

EXAMPLE 1.2—If the seed analyst in the foregoing example had tested 1,000 seeds instead of 100 and had found only 8% ungerminated, what statement might he have made with 99:1 odds of being right? Ans. Between 90% and 94% will germinate.

EXAMPLE 1.3—An investigator making a nation-wide survey interviewed 115 women over 40 years of age from the lower middle economic level in rural areas of middle western states. Forty-six of them had listened to a certain radio program three or more times during the preceding month. Assuming random sampling, what statement can be made about the population fraction of such women listening, using the 95% confidence interval? Ans. Approximately, between 30.4% and 49.6% listen. You will have to interpolate these values in the table.

EXAMPLE 1.4—A sampler of public opinion asked 50 men to express their preferences between candidates A and B. Twenty preferred A. Assuming random sampling from a population of 5,000, the sampler stated that between 1,150 and 2,950 men of the population preferred A. What was the probability of his having been led into error by his sampling? Ans. 0.01.

EXAMPLE 1.5—In a random sample from a uniform population, 86% expressed approval of a certain agricultural program. On the basis of this sample the statistician asserted that unless a one-in-twenty chance had come off the percentage of those favorable in the sampled population was between 81% and 90%. What was the size of the sample? Ans. 250.

EXAMPLE 1.6—If you guess that in a certain uniform stratum about half of the housewives own a specified appliance, and if you wish to survey a sample that will, at the 95% confidence level, yield an estimate differing by not more than 6 from the correct percentage, about how large a sample must you take? Ans. 250.

EXAMPLE 1.7—If you draw a sample properly, then make a statement about the population ratio, the statement will be either right or wrong but you will not know which. Do you see clearly how chance enters into the making of the statement?

EXAMPLE 1.8—After the survey of a certain underprivileged community it was reported (14) that among 978 children less than age 16, 54.4% were girls. Does the 95% confidence interval include a reasonable sex ratio for this group?

EXAMPLE 1.9—The F_2 plants of a certain cross in maize segregated in the ratio 118 normal green to 71 chlorophyll deficient. What common genetic ratios might be appropriate?

1.4—Sampling: definitions and discussions. Size of sample. I have rushed you along through the first three sections, thinking you might enjoy a quick glimpse into the ideas and utility of statistics. Some points need considerable clarification. A few of them will be discussed in this section, while others will come up later.

The heading of this chapter is "Sampling of Attributes." So far we have considered such diverse attributes as shortage of farm labor, germination of seed, approval of an agricultural program, and chlorophyll deficiency in plants. The possession or lack of an attribute distinguishes the two classes of individuals making up the populations we have been studying. The record of the sampling consists of the numbers of individuals found to have or lack the attribute under investigation. I call these records *enumeration data* in distinction from records of measurements of such variables as height or age. The theory and methods for *measurement data* will be developed in chapters 2 and 3.

One must learn to distinguish clearly between those elements which are certain and those which are matters of chance. The possession of an attribute by an individual is a fact that may be certainly determined, but the selection of this individual in a sample is a chance event. Often heard is some such inaccurate reporting as, "The probability that an Iowa family owns a car is 90%." An Iowa family either owns a car or doesn't, the accurate phraseology being, "The probability is 90% that a randomly selected Iowa family shall be found to own a car." It is the selection of the family that is the chance event, not the ownership of the car. This distinction is difficult in statements about the confidence interval. The population ratio either does or does not lie within any stated interval, irrespective of the sampling. Furthermore, there is no uncertainty about reading the table. The chance lies in the drawing of the sample. Selection of the last member of the sample ends all uncertainty about the truth or falsity of the inference that will be made, despite the fact that the sampler may never learn whether his statement has been right or wrong. The probability involved is the probability that the sample drawn lead to a correct conclusion. If you make a practice of sampling, and if for each sample you assert that the population ratio lies within the 0.95 confidence interval for that sample, an average of 95 per hundred of your assertions will be right. Note that this theory rests solidly on the randomness of the sampling.

The term “unbiased” requires some further comment. An unbiased sample implies an unbiased estimate of the population ratio, an estimate that tends toward the true value as the size of the sampling increases. This feature of random sampling is reflected in the narrowing confidence interval that accompanies expansion of sample size. In sampling from certain symmetrical populations, it may be said that estimates made from unbiased samples are as likely to be in excess of the population value as in defect. It is clear that to say “unbiased” is not to imply “without sampling variation.” While working the examples you were doubtless impressed by the fact that an estimate from a small sample, even though unbiased, may be very inaccurate. The point is that in unbiased samples accuracy increases with size, but the same cannot be said about biased sampling.

Randomness in sampling is perhaps never quite attained in practice. It is nevertheless the mathematical model on which much statistical theory rests, and since the theory must be used in drawing conclusions from work-a-day samplings, it is to the interest of the investigator to approximate, as closely as feasible, the ideal conditions. The better the approximation the more nearly correct will be the inferences drawn.

There are two common devices used to approach randomness of sampling; the population itself may be thoroughly mixed, or else some scheme may be found for leaving to chance the choice of the individuals taken. Often a combination is used, as in the sampling of farmers from the assessors' books: apparently the listing of the names was random, but the precaution was taken of using every fifth. In the next section you will assume complete randomness in the lists and take the sample consecutively. But it is a matter of experience that most lists, as well as most natural populations, have groupings into families, communities, social and economic levels, etc., which require much care if the operation of chance is to be insured. Several available methods will be discussed later.

The question most often asked by experimenters and other designers of samplings is, “How large a sample must I take?” The answers that may be given in various circumstances are not easy and always involve chance variation. Nevertheless, at appropriate places in this text (see Index) attempts will be made to clarify the problems involved. You doubtless noticed the method of attack proposed in example 1.6, where the confidence interval and a guess about the population percentage were proposed as data. Perhaps you noticed, also, that the method doesn't work so well in the upper and left hand portions of the table; not only do the confidence limits change rapidly with the unknown X , but they are not equally spaced above and below this fraction. Despite these shortcomings of the table, answers can usually be got from it sufficiently good for practical purposes. It is futile to attempt exact estimates of sample size because unpredictable sampling variation will upset the best laid plans. At the present stage my advice is to read the table as accurately as you may, then take a somewhat larger sample if you can afford it.

1.5—An exercise in sampling. Whenever I draw samples from a known population I am amazed at the capricious way in which the items turn up. It is a salutary discipline for an experimenter or enumerator occasionally to observe the laws of chance in action lest he become too confident of his professional samplings. Drawing marbles from an urn or cards from a pack, tossing coins and throwing dice are the classical methods of doing this. Heretofore I have asked my students to collect their samples by drawing marked or numbered beans from a bag, but the more modern method is to use a table of random numbers such as table 1.2. This table contains 10,000 digits jumbled together presumably in random fashion, the 5 x 5 blocks serving merely to facilitate reading. There are 100 rows and 100 columns, each numbered from 00 to 99. Instead of examining peas to learn if they are wrinkled or smooth, or instead of asking people to vote "Yes" or "No," you can draw a random sample of digits from this table and observe whether each is odd or even. Since the digits are supposed to be thoroughly mixed, with no particular order or groupings, any sequence of them may be considered random; so, begin at any point in the table and count the number of odd digits in a sample of some predetermined size. For example, suppose you decide on 50 digits as the sample size, and elect to start at row 31, column 17. You find

752 54450 19031 . . .

Read on to the end of this row, then return to the beginning of row 32 and continue till 50 digits have been surveyed, ending with

33851 44

In this particular sample you will count 23 odd digits; that is, 46% of the sample of 50 are odd.

Three ways of introducing elements of chance into your sampling will decrease the likelihood of duplicating the drawings of others. (i) Start at a randomly selected point for each sample. A good way to do this is to place your pencil aimlessly on some digit in the table, then use this and the subsequent three digits to fix the initial point of your first sample. As an example, suppose your pencil point picks out 2 in row 80, column 84. Let this with the three digits following, 2061, specify row 20, column 61 as the point to begin sample number one. For the second sample, start at a point fixed by the first four (or last four or any other convenient four) digits of the first one, and so on. (ii) Vary the direction in which you traverse the table; to the right or left, up or down, or diagonally. (iii) Vary the sample size, choosing in some random manner the various sizes, 10, 15, . . . 250, heading table 1.1 (1,000 is unnecessarily large).

For each sample make a record of the following facts: (i) initial point, (ii) direction taken in counting, (iii) sample size, and (iv) the number of odd digits. If you can spare the time, draw 100 or more samples, carefully preserving the records, in the order made, for future use.

TABLE 1.2
TEN THOUSAND RANDOMLY ASSORTED DIGITS

	00-04	05-09	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49
00	54463	22662	65905	70639	79365	67382	29085	69831	47058	08186
01	15389	85205	18850	39226	42249	90669	96325	23248	60933	26927
02	85941	40756	82414	02015	13858	78030	16269	65978	01385	15345
03	61149	69440	11286	88218	58925	03638	52862	62733	33451	77455
04	05219	81619	10651	67079	92511	59888	84502	72095	83463	75577
05	41417	98326	87719	92294	46614	50948	64886	20002	97365	30976
06	28357	94070	20652	35774	16249	75019	21145	05217	47286	76305
07	17783	00015	10806	83091	91530	36466	39981	62481	49177	75779
08	40950	84820	29881	85966	62800	70326	84740	62660	77379	90279
09	82995	64157	66164	41180	10089	41757	78258	96488	88629	37231
10	96754	17676	55659	44105	47361	34833	86679	23930	53249	27083
11	34357	88040	53364	71726	45690	66334	60332	22554	90600	71113
12	06318	37403	49927	57715	50423	67372	63116	48888	21505	80182
13	62111	52820	07243	79931	89292	84767	85693	73947	22278	11551
14	47534	09243	67879	00544	23410	12740	02540	54440	32949	13491
15	98614	75993	84460	62846	59844	14922	48730	73443	48167	34770
16	24856	03648	44898	09351	98795	18644	39765	71058	90368	44104
17	96887	12479	80621	66223	86085	78285	02432	53342	42846	94771
18	90801	21472	42815	77408	37390	76766	52615	32141	30268	18106
19	55165	77312	83666	36028	28420	70219	81369	41943	47366	41067
20	75884	12952	84318	95108	72305	64620	91318	89872	45375	85436
21	16777	37116	58550	42958	21460	43910	01175	87894	81378	10620
22	46230	43877	80207	88877	89380	32992	91380	03164	98656	59337
23	42902	66892	46134	01432	94710	23474	20423	60137	60609	13119
24	81007	00333	39693	28039	10154	95425	39220	19774	31782	49037
25	68089	01122	51111	72373	06902	74373	96199	97017	41273	21546
26	20411	67081	89950	16944	93054	87687	96693	87236	77054	33848
27	58212	13160	06468	15718	82627	76999	05999	58680	96739	63700
28	70577	42866	24969	61210	76046	67699	42054	12696	93758	03283
29	94522	74358	71659	62038	79643	79169	44741	05437	39038	13163
30	42626	86819	85651	88678	17401	03252	99547	32404	17918	62880
31	16051	33763	57194	16752	54450	19031	58580	47629	54132	60631
32	08244	27647	33851	44705	94211	46716	11738	55784	95374	72655
33	59497	04392	09419	89964	51211	04894	72882	17805	21896	83864
34	97155	13428	40293	09985	58434	01412	69124	82171	59058	82859
35	98409	66162	95763	47420	20792	61527	20441	39435	11859	41567
36	45476	84882	65109	96597	25930	66790	65706	61203	53634	22557
37	89300	69700	50741	30329	11658	23166	05400	66669	48708	88887
38	50051	95137	91631	66315	91428	12275	24816	68091	71710	33258
39	31753	85178	31310	89642	98364	02306	24617	09609	83942	22716
40	79152	53829	77250	20190	56535	18760	69942	77448	33278	48805
41	44560	38750	83635	56540	64900	42912	13953	79149	18710	68618
42	68328	83378	63369	71381	39564	05615	42451	64559	97501	65747
43	46939	38689	58625	08342	30459	85863	20781	09284	26333	91777
44	83544	86141	15707	96256	23068	13782	08467	89469	93842	55349
45	91621	00881	04900	54224	46177	55309	17852	27491	89415	23466
46	91896	67126	04151	03795	59077	11848	12630	98375	52068	60142
47	55751	62515	21108	80830	02263	29303	37204	96926	30506	09808
48	85156	87689	95493	88842	00664	55017	55539	17771	69448	87530
49	07521	56898	12236	60277	39102	62315	12239	07105	11844	01117

TABLE 1.2—(Continued)

	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85-89	90-94	95-99
00	59391	58030	52098	82718	87024	82848	04190	96574	90464	29065
01	99567	76364	77204	04615	27062	96621	43918	01896	83991	51141
02	10363	97518	51400	25670	98342	61891	27101	37855	06235	33316
03	86859	19558	64432	16706	99612	59798	32803	67708	15297	28612
04	11258	24591	36863	55368	31721	94335	34936	02566	80972	08188
05	95068	88628	35911	14530	33020	80428	39936	31855	34334	64865
06	54463	47237	73800	91017	36239	71824	83671	39892	60518	37092
07	16874	62677	57412	13215	31389	62233	80827	73917	82802	84420
08	92494	63157	76593	91316	03505	72389	96363	52887	01087	66091
09	15669	56689	35682	40844	53256	81872	35213	09840	34471	74441
10	99116	75486	84989	23476	52967	67104	39495	39100	17217	74073
11	15696	10703	65178	90637	63110	17622	53988	71087	84148	11670
12	97720	15369	51269	69620	03388	13699	33423	67453	43269	56720
13	11666	13841	71681	98000	35979	39719	81899	07449	47985	46967
14	71628	73130	78783	75691	41632	09847	61547	18707	85489	69944
15	40501	51089	99943	91843	41995	88931	73631	69361	05375	15417
16	22518	55576	98215	82068	10798	86211	36584	67466	69373	40054
17	75112	30485	62173	02132	14878	92879	22281	16783	86352	00077
18	80327	02671	98191	84342	90813	49268	95441	15496	20168	09271
19	60251	45548	02146	05597	48228	81366	34598	72856	66762	17002
20	57430	82270	10421	05540	43648	75888	66049	21511	47676	33444
21	73528	39559	34434	88596	54086	71693	43132	14414	79949	85193
22	25991	65959	70769	64721	86413	33475	42740	06175	82758	66248
23	78388	16638	09134	59880	63806	48472	39318	35434	24057	74739
24	12477	09965	96657	57994	59439	76330	24596	77515	09577	91871
25	83266	32883	42451	15579	38155	29793	40914	65990	16255	17777
26	76970	80876	10237	39515	79152	74798	39357	09054	73579	92359
27	37074	65198	44785	68624	98336	84481	97610	78735	46703	98265
28	83712	06514	30101	78295	54656	85417	43189	60048	72781	72606
29	20287	56862	69727	94443	64936	08366	27227	05158	50326	59566
30	74261	32592	86538	27041	65172	85532	07571	80609	39285	65340
31	64081	49863	08478	96001	18888	14810	70545	89755	59064	07210
32	05617	75818	47750	67814	29575	10526	66192	44464	27058	40467
33	26793	74951	95466	74307	13330	42664	85515	20632	05497	33625
34	65988	72850	48737	54719	52056	01596	03845	35067	03134	70322
35	27366	42271	44300	73399	21105	03280	73457	43093	05192	48657
36	56760	10909	98147	34736	33863	95256	12731	66598	50771	83665
37	72880	43338	93643	58904	59543	23943	11231	83268	65938	81581
38	77888	38100	03062	58103	47961	83841	25878	23746	55903	44115
39	28440	07819	21580	51459	47971	29882	13990	29226	23608	15873
40	63525	94441	77033	12147	51054	49955	58312	76923	96071	05813
41	47606	93410	16359	89033	89696	47231	64498	31776	05383	39902
42	52669	45030	96279	14709	52372	87832	02735	50803	72744	88208
43	16738	60159	07425	62369	07515	82721	37875	71153	21315	00132
44	59348	11695	45751	15865	74739	05572	32688	20271	65128	14551
45	12900	71775	29845	60774	94924	21810	38636	33717	67598	82521
46	75086	23537	49939	33595	13484	97588	28617	17979	70749	35234
47	99495	51434	29181	09993	38190	42553	68922	52125	91077	40197
48	26075	31671	45386	36583	93459	48599	52022	41330	60651	91321
49	13636	93596	23377	51133	95126	61496	42474	45141	46660	42338

TABLE 1.2—(Continued)

	00-04	05-09	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49
50	64249	63664	39652	40646	97306	31741	07294	84149	46797	82487
51	26538	44249	04050	48174	65570	44072	40192	51153	11397	58212
52	05845	00512	78630	55328	18116	69296	91705	86224	29503	57071
53	74897	68373	67359	51014	33510	83048	17056	72506	82949	54600
54	20872	54570	35017	88132	25730	22626	86723	91691	13191	77212
55	31432	96156	89177	75541	81355	24480	77243	76690	42507	84362
56	66890	61505	01240	00660	05873	13568	76082	79172	57913	93448
57	48194	57790	79970	33106	86904	48119	52503	24130	72824	21627
58	11303	87118	81471	52936	08555	28420	49416	44448	04269	27029
59	54374	57325	16947	45356	78371	10563	97191	53798	12693	27928
60	64852	34421	61046	90849	13966	39810	42699	21753	76192	10508
61	16309	20384	09491	91588	97720	89846	30376	76970	23063	35894
62	42587	37065	24526	72602	57589	98131	37292	05967	26002	51945
63	40177	98590	97161	41682	84533	67588	62036	49967	01990	72308
64	82309	76128	93965	26743	24141	04838	40254	26065	07938	76236
65	79788	68243	59732	04257	27084	14743	17520	95401	55811	76099
66	40538	79000	89559	25026	42274	23489	34502	75508	06059	86682
67	64016	73598	18609	73150	62463	33102	45205	87440	96767	76702
68	49767	12691	17903	93871	99721	79109	09425	26904	07419	76013
69	76974	55108	29795	08404	82684	00497	51126	79935	57450	55671
70	23854	08480	85983	96025	50117	64610	99425	62291	86943	21541
71	68973	70551	25098	78033	98573	79848	31778	29555	61446	23037
72	36444	93600	65350	14971	25325	00427	52073	64280	18847	24768
73	03003	87800	07391	11594	21196	00781	32550	57158	58887	73041
74	17540	26188	36647	78386	04558	61463	57842	90382	77019	24210
75	38916	55809	47982	41968	69760	79422	80154	91486	19180	15100
76	64288	19843	69122	42502	48508	28820	59933	72998	99942	10515
77	86809	51564	38040	39418	49915	19000	58050	16899	79952	57849
78	99800	99566	14742	05028	30033	94889	53381	23656	75787	59223
79	92345	31890	95712	08279	91794	94068	49337	88674	35355	12267
80	90363	65162	32245	82279	79256	80834	06088	99462	56705	06118
81	64437	32242	48431	04835	39070	59702	31508	60935	22390	52246
82	91714	53662	28373	34333	55791	74758	51144	18827	10704	76803
83	20902	17646	31391	31459	33315	03444	55743	74701	58851	27427
84	12217	86007	70371	52281	14510	76094	96579	54853	78339	20839
85	45177	02863	42307	53571	22532	74921	17735	42201	80540	54721
86	28325	90814	08804	52746	47913	54577	47525	77705	95330	21866
87	29019	28776	56116	54791	64604	08815	46049	71186	34650	14994
88	84979	81353	56219	67062	26146	82567	33122	14124	46240	92973
89	50371	26347	48513	63915	11158	25563	91915	18431	92978	11591
90	53422	06825	69711	67950	64716	18003	49581	45378	99878	61130
91	67453	35651	89316	41620	32048	70225	47597	33137	31443	51445
92	07294	85353	74819	23445	68237	07202	99515	62282	53809	26685
93	79544	00302	45338	16015	66613	88968	14595	63836	77716	79596
94	64144	85442	82060	46471	24162	39500	87351	36637	42833	71875
95	90919	11883	58318	00042	52402	28210	34075	33272	00840	73268
96	06670	57353	86275	92276	77591	46924	60839	55437	03183	13191
97	36634	93976	52062	83678	41256	60948	18685	48992	19462	96062
98	75101	72891	85745	67106	26010	62107	60885	37503	55461	71213
99	05112	71222	72654	51583	05228	62056	57390	42746	39272	96659

TABLE 1.2—(Continued)

	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85-89	90-94	95-99
50	32847	31282	03345	89593	69214	70381	78285	20054	91018	16742
51	16916	00041	30236	55023	14253	76582	12092	86533	92426	37655
52	66176	34047	21005	27137	03191	48970	64625	22394	39622	79085
53	46299	13335	12180	16861	38043	59292	62675	63631	37020	78195
54	22847	47839	45385	23289	47526	54098	45683	55849	51575	64689
55	41851	54160	92320	69936	34803	92479	33399	71160	64777	83378
56	28444	59497	91586	95917	68553	28639	06455	34174	11130	91994
57	47520	62378	98855	83174	13088	16561	68559	26679	06238	51254
58	34978	63271	13142	82681	05271	08822	06490	44984	49307	62717
59	37404	80416	69035	92980	49486	74378	75610	74976	70056	15478
60	32400	65482	52099	53676	74648	94148	65095	69597	52771	71551
61	89262	86332	51718	70663	11623	29834	79820	73002	84886	03591
62	86866	09127	98021	03871	27789	58444	44832	36505	40672	30180
63	90814	14833	08759	74645	05046	94056	99094	65091	32663	73040
64	19192	82756	20553	58446	55376	88914	75096	26119	83898	43816
65	77585	52593	56612	95766	10019	29531	73064	20953	53523	58136
66	23757	16364	05096	03192	62386	45389	85332	18877	55710	96459
67	45989	96257	23850	26216	23309	21526	07425	50254	19455	29315
68	92970	94243	07316	41467	64837	52406	25225	51553	31220	14032
69	74346	59596	40088	98176	17896	86900	20249	77753	19099	48885
70	87646	41309	27636	45153	29988	94770	07255	70908	05340	99751
71	50099	71038	45146	06146	55211	99429	43169	66259	97786	59180
72	10127	46900	64984	75348	04115	33624	68774	60013	35515	62556
73	67995	81977	18984	64091	02785	27762	42529	97144	80407	64524
74	26304	80217	84934	82657	69291	35397	98714	55104	08187	48109
75	81994	41070	56642	64091	31229	02595	13513	45148	78722	30144
76	59537	34662	79631	89403	65212	09975	06118	86197	58208	16162
77	51228	10937	62396	81460	47331	91403	95007	06047	16846	64809
78	31089	37995	29577	07828	42272	54016	21950	86192	99046	84864
79	38207	97938	93459	75174	79460	55436	57206	87644	21296	43395
80	88666	31142	09474	89712	63153	62333	42212	06140	42594	43671
81	53365	56134	67582	92557	89520	33452	05134	70628	27612	33738
82	89807	74530	38004	90102	11693	90257	05500	79920	62700	43325
83	18682	81038	85662	90915	91631	22223	91588	80774	07716	12548
84	63571	32579	63942	25371	09234	94592	98475	76884	37635	33608
85	68927	56492	67799	95398	77642	54913	91853	08424	81450	76229
86	56401	63186	39389	88798	31356	89235	97036	32341	33292	73757
87	24333	95603	02359	72942	46287	95382	08452	62862	97869	71775
88	17025	84202	95199	62272	06366	16175	97577	99304	41587	03686
89	02804	08253	52133	20224	68034	50865	57868	22343	55111	03607
90	08298	03879	20995	19850	73090	13191	18963	82244	78479	99121
91	59883	01785	82403	96062	03785	03488	12970	64896	38336	30030
92	46982	06682	62864	91837	74021	89094	39952	64158	79614	78235
93	31121	47266	07661	02051	67599	24471	69843	83696	71402	76287
94	97867	56641	63416	17577	30161	87320	37752	73276	48969	41915
95	57364	86746	08415	14621	49430	22311	15836	72492	49372	44103
96	09559	26263	69511	28064	75999	44540	13337	10918	79846	54809
97	53873	55571	00608	42661	91332	63956	74087	59008	47493	99581
98	35531	19162	86406	05299	77511	24311	57257	22826	77555	05941
99	28229	88629	25695	94932	30721	16197	78742	34974	97528	45447

You have now done experimentally what the mathematical statistician does theoretically when he studies the distribution of samples drawn at random from a specified population, and you can verify the results which he derives from his equations. First count the number of samples whose percentage of odd digits is greater than 50.6%, which is the exact fraction of odds in the population of 10,000. Approximately half of your percentages should be above this population value and half below. Do you find much sampling deviation from half-and-half? If you have 100 samples, you need not be surprised if 45 percentages are on one side of 50.6% and 55 on the other. The second point to observe is a characteristic of nearly all sampling which should be rather obvious in yours; that is, the tendency of the sample percentages to concentrate near the population 50.6%. How many of yours were between say, 45% and 55%? This central tendency is an idea which should counterbalance the perhaps discouraging fact that some samples are notably divergent.

Next, you have evidence of the unbiased nature of the sampling. Look over the percentages in all your samples of 10 and note the greatest deviation from 50.6%. Do the same with your samples of 15, 20, etc. The larger samples should not have such pronounced deviations from 50.6% as are notable in some of the smaller. This verifies the fact that as the sample size increases its average tends towards the population value; not invariably, indeed, but with constantly increasing probability of small deviations.

Finally, you can verify the theory of the confidence interval. From table 1.1, write down the 95% interval for each of your samples. Thus, for the sample of 50 discussed above, 23 of the digits were odd. This leads to the statement that the population percentage lies between 32% and 61%, a correct conclusion because we know that 50.6% of the digits in the table are odd. But suppose one of your samples of 250 had started at row 85, column 23, you would have counted, moving down the successive columns, only 101 or 40.4% odd, and would have asserted that the true value is between 34% and 46%. You would have been wrong despite the fact that this sample is randomly drawn from the same population as the others. It merely happens to be unusually divergent. You should find about five samples in a hundred leading you to incorrect statements, but there will be no occasion for surprise if only three, or as many as seven, turn up. If you now examine the intervals for $P = 0.99$, you may expect (allowing as usual for sampling variation) about 99 per 100 of these to include the true ratio of 50.6%. I hope your results may be sufficiently concordant with theory to give you confidence in it. Of one thing I am sure—you are more aware of the vagaries of sampling than you were before, and that is one of the objects of the experiment. Another lesson to be learned is that only broad confidence intervals can be based on small samples, and that even so the inference can be wrong. Also, one finds that refinement of the statements increases but slowly with sample size.

Let me emphasize the point that failure to make correct inferences in a small portion of samples is not a fault that can be remedied but is inevitably bound up in the sampling procedure. Fallibility is the very nature of such evidence. The sampler can only take available precautions, then prepare himself for his share of mistakes. In this he is not alone. The journalist, the judge, the banker, the weather forecaster—these and most other people are subject to the laws of chance, and each makes his own quota of wrong guesses. The statistician has this advantage—he can, in favorable circumstances, know exactly his likelihood of error.

EXAMPLE 1.10—If you wish to learn by sampling what portion of radio owners in a selected region listen to a certain program, would you get an unbiased sample if you conducted the interviews by telephone? Would all radio owners have the same chance of being interviewed?

EXAMPLE 1.11—A sampler of public opinion estimates from a sample the number of voters in a state favoring a certain candidate for governor. Assuming that his estimate was very close to the population value at the time the survey was made, might the ballot on election day be quite different?

EXAMPLE 1.12—Why does a mailed questionnaire often result in a biased estimate?

EXAMPLE 1.13—Combine all your samples of digits, irrespective of size, into one. The correct percentage of odd digits can be got in two ways: (i) add all the counts of odd digits and divide by the total digits counted; (ii) compute the weighted average of the percentages of the various sized samples, using the formula,

$$\text{weighted average} = \frac{W_1 p_1 + W_2 p_2 + W_3 p_3 + \dots}{W_1 + W_2 + W_3 + \dots}$$

where p_1 is the percentage of odds in the samples of 10 and W_1 is the total number of digits counted in these samples. p_2 and W_2 are defined in the same way for samples of 15, etc.

EXAMPLE 1.14—If you have 100 swine to be distributed at random into 10 lots, how can the allotment be made by means of table 1.2?

1.6—Hypotheses about populations. In contrast with the foregoing, the investigator often has in mind a definite hypothesis about the population ratio, the purpose of the sampling being to get evidence concerning his hypothesis. Thus, a geneticist studying heredity in the tomato had reason to believe that, in the F_2 generation, segregation of fruits with red flesh and yellow flesh would be in the ratio, 3:1. Counting the number of fruits with red flesh in a sample of 400 he found 310 instead of the expected 300. With your experience of sampling variation, would you accept this as verification or refutation of the hypothesis? Again, the sampler of public opinion foresees a close election in some critical state. He sets up the hypothesis of an equal number of votes for the two dominant parties, then asks 10,000 voters their preferences, the resulting division being 5,100 to 4,900. Shall he say that this is or is not convincing evidence about his hypothesis of a tie vote?

To answer such questions two data are needed, a measure of the deviation of the sample from the hypothetical population ratio, and a

means of judging whether or not this measure is an amount commonly expected in sampling or, on the contrary, is so great as to throw doubt upon the hypothesis. Both data were furnished by Karl Pearson in 1899 (11). He devised an index of dispersion denoted by χ^2 (chi-square), and provided a table of its probable occurrence in sampling.* Let us first examine the index.

1.7—Chi-square, an index of dispersion. Naturally, the deviations of the observed numbers from those specified by the hypothesis form the basis of the index. In the election sampling of the previous section, the numbers expected on the assumption of an equally divided vote are each 5,000. The deviations, then, are

$$5,100 - 5,000 = 100$$

and

$$4,900 - 5,000 = -100,$$

the sum of the two being zero. The value of chi-square is given by,

$$\chi^2 = \frac{(100)^2}{5,000} + \frac{(-100)^2}{5,000} = 4$$

Each deviation is squared, each square is divided by the expected number, and the results are added. The expected numbers appear in the denominators in order to introduce sample size into the quantity—it is the relative size of the deviation which is important. The squaring may puzzle you. You will find this a very common practice in statistics. Incidentally, it makes the sign of the deviation unimportant since the square of a negative number is the same as that of the corresponding positive. It is plain that chi-square would be zero if the sample totals were the same as those expected, and that it will increase with increasing deviation from expected. In that respect, it is a reasonable measure of variation. But it is not at all plain whether the chi-square value of 4 is to be considered large, medium, or small. To furnish you a basis for comparison is our next aim. Pearson founded his judgment on the solution and tabulation of a somewhat elaborate equation, but for non-mathematicians the customary way of going about the matter is to set up an experiment. Before doing that, a couple of formulas will be given, together with a few examples to help fix them in mind.

1.8—Symbolical statement of chi-square. It is convenient to represent by X_1 and X_2 the sample counts of individuals who do and do not possess the attribute being investigated, the corresponding expected numbers being m_1 and m_2 . The two deviations, then, are $X_1 - m_1$ and $X_2 - m_2$, so that chi-square is given by the formula,

$$\chi^2 = (X_1 - m_1)^2/m_1 + (X_2 - m_2)^2/m_2$$

* G. U. Yule and R. A. Fisher discovered an error in certain uses of the table and corrected it in 1922 (17) (4).

The formula may be condensed to the more easily remembered as well as more general one,

$$\chi^2 = S(X - m)^2/m,$$

where S denotes summation. In words, "Chi-square is the sum of such ratios as

$$(\text{deviation square})/(\text{expected number})"$$

As will appear later, there may be more than two of these ratios in the summation.

Let us apply the formula to the counts of red and yellow tomatoes in section 1.6. There, $X_1 = 310$, $X_2 = 400 - 310 = 90$, $m_1 = \frac{3}{4}$ of 400 = 300, and $m_2 = \frac{1}{4}$ of 400 = 100. Whence,

$$\chi^2 = \frac{(310 - 300)^2}{300} + \frac{(90 - 100)^2}{100} = 1.33$$

It is apparent that the hypothesis set up about the population is the sole criterion for dividing the sample into expected numbers. Only if the investigator wishes to test some hypothesis will he be interested in calculating chi-square.

EXAMPLE 1.15—In a certain cross of two varieties of peas genetic theory led the investigator to expect half wrinkled seeds and half smooth. In a sample of 800 seeds examined he found 440 wrinkled. What is the value of chi-square? Ans. 8.

EXAMPLE 1.16—If the count in the foregoing example had been 220 wrinkled in a total of 400, would chi-square also be half its original value?

EXAMPLE 1.17—In the text example about tomatoes the deviation from expected was 10. If this same deviation had occurred in a sample of twice the size (that is, in a sample of 800), what would have been the value of chi-square? Ans. 0.67, half the original value.

EXAMPLE 1.18—In a political campaign a candidate claimed that 60% of the electorate would vote for him. A sampler of public opinion asked 1,000 registered voters if they expected to vote for this candidate and 55% of them said, "Yes." Calculate the value of chi-square on the assumption that the candidate's claim was correct. Ans. 10.42.

EXAMPLE 1.19—In the foregoing example, what statement about the electorate may the sampler make based on the 99% confidence interval? Ans. Between 51% and 59% will vote for the candidate.

1.9—An experiment in sampling chi-square. You have now had some practice in the calculation of chi-square. It is to be hoped that it appeals to you as a reasonable and informative measure of the deviation of the observed sample numbers from those expected under the hypothesis set up. But it has another function—it enables one to judge whether the sample ratio itself departs much or little from the hypothetical population value. To use it for that purpose we must answer the question already proposed: what values of chi-square are to be considered as indicating unusual deviation, and what as ordinary sampling variation? Our experimental method of answering the question will be to calculate chi-square

for each of the samples you drew from the table of random numbers, then observe what values of chi-square spring from the more unusual samples. If a large number of samples of various sizes have been drawn, and if the value of chi-square is computed from each, the distribution of chi-square may be mapped.

To illustrate the calculation of chi-square for one of your samples suppose the sample size is $n = 100$; then, since 50.6% of the tabular digits are odd, $m_1 = (0.506)(100) = 50.6$ and $m_2 = 49.4$. If you counted $X_1 = 44$ odd digits in this sample, the value of chi-square is

$$\chi^2 = \frac{(44 - 50.6)^2}{50.6} + \frac{(56 - 49.4)^2}{49.4} = 1.74$$

In case there are several samples of this same size, calculating time can be saved by rearranging the formula. Remembering that the two deviations are the same in absolute value (that is, disregarding sign), this absolute value being 6.6 for each deviation, chi-square can be written

$$\begin{aligned}\chi^2 &= (X - m)^2(1/m_1 + 1/m_2) \\ &= (6.6)^2(1/50.6 + 1/49.4) \\ &= (43.56)(0.019763 + 0.020243) \\ &= (43.56)(0.040006) = 1.74\end{aligned}$$

as before. The advantage is that once the sum of the reciprocals of the expected numbers has been calculated, it can be used again and again so long as the sample size and the population ratio are unchanged: just square the deviation from expected and multiply by 0.040006. Of course, an appropriate factor must be calculated for each sample size.

Those who do not have adequate computing facilities (table of squares, slide rule, calculating machine) may prefer to sacrifice some precision to save a good deal in the time required for the experiment just outlined. If they assume 50% of the tabular digits to be odd, then $m_1 = m_2 = n/2$, an integer in most of the samples. For $n = 100$, the formula above then becomes

$$\chi^2 = (0.04)(X - 50)^2$$

The effect on the frequency distribution may not be noticed, but the individual values of chi-square in large samples are obviously different.

Proceed now to calculate chi-square for each of your samples. The results may be summarized conveniently, especially if there are a hundred or more of them, in tables and a graph to be described in the next sections.

1.10—The frequency distribution. Table 1.3 is a typical *frequency distribution*, or merely, *distribution*. *Class intervals* (0.00 to 0.49, etc.) of convenient size are chosen, then the values falling into each are counted (in this table the values are those of chi-square). The counts are called *class frequencies*. You may find it more convenient to halve the number of classes for your distribution, making the first extend from zero to 0.99.

TABLE 1.3
DISTRIBUTION OF 230 VALUES OF CHI-SQUARE CALCULATED FROM SAMPLES DRAWN
FROM TABLE 1.2
Sample sizes—10, 15, 20, 30, 50, 100, and 250

Class Interval	Frequency	Class Interval	Frequency
0.00-0.49	116	6.00- 6.49	0
0.50-0.99	39	6.50- 6.99	1
1.00-1.49	18	7.00- 7.49	0
1.50-1.99	22	7.50- 7.99	0
2.00-2.49	12	8.00- 8.49	0
2.50-2.99	5	8.50- 8.99	1
3.00-3.49	5	9.00- 9.49	0
3.50-3.99	6	9.50- 9.99	0
4.00-4.49	1	10.00-10.49	1
4.50-4.99	2	10.50-10.99	0
5.00-5.49	0	11.00-11.49	1
5.50-5.99	0		

For more detailed information about the structure of statistical tables, see reference (15).

Observe the concentration of sample chi-squares in the smallest class, practically half of them being less than 0.5. Attention has already been called to the fact that in the process of sampling, small deviations are predominant. (This tendency of sample ratios to deviate little from the population value is the foundation of confidence in sampling.) But taking a less optimistic view, one must not overlook the samples with large deviations and chi-squares. The possibility of getting such as these makes for caution in drawing conclusions. In this sampling exercise we know the population ratio and are not led astray by discrepant samples, but in actual investigations, where the hypothesis set up is not known to be the right one, a large value of chi-square constitutes a dilemma. Shall we say that it denotes only an unusual sample from the hypothetical population, or shall we conclude that the hypothesis misrepresents the true population ratio? Statistical theory contains no certain answer. Instead, it furnishes an evaluation of the probability of possible sample deviations from the hypothetical population. (If chi-square is large, the investigator is warned that the sample is an improbable one under his hypothesis.) This is evidence to be added to that which he already possesses, all of it being the basis for his decisions. A more exact determination of probability will be explained in the second section to follow.

1.11—Graphical representation. You will no doubt wish to represent your sample chi-squares in a graph such as figure 1.1. On the horizontal axis mark off intervals corresponding to the classes of the frequency distribution. Upon each construct a rectangle with altitude proportional to the corresponding class frequency, using the scale of frequencies at the left. This kind of graphical representation of the frequency distribution

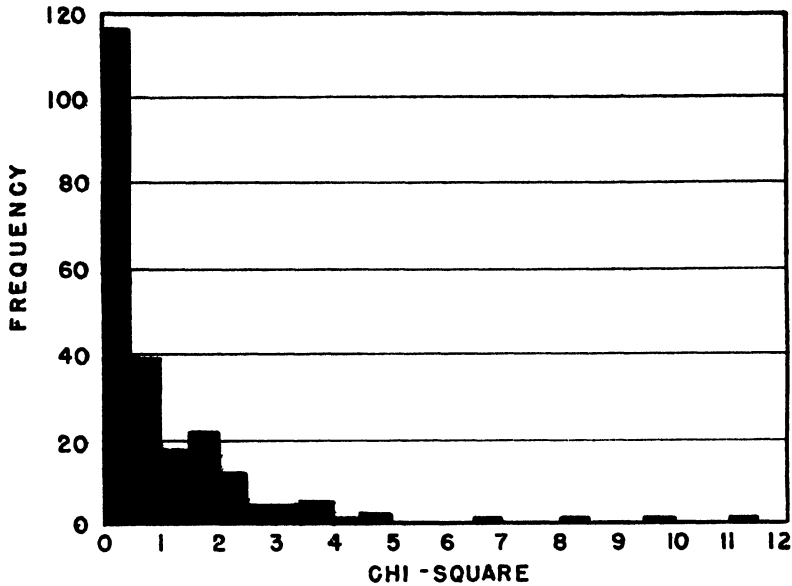


FIG. 1.1—Histogram representing frequency distribution of the 230 sample values of chi-square in table 1.3.

is known as a *histogram*. The graph serves to emphasize both the concentration of small chi-squares at the left and the comparatively large size of the few at the right. Reference (1) contains excellent instructions for the making of graphs.

1.12—Frequency distribution with unequal class intervals. Two features of our chi-square distribution are yet to be examined: (i) how does it compare with the theoretical distribution, and (ii) how may we evaluate more exactly the probabilities of various chi-square sizes? Both features become evident if the class intervals are rearranged as in table 1.4, column 1. Very narrow at the start, these theoretically determined intervals increase in width so as to make the chi-square distribution symmetrical (5,6) as in column 4. Furthermore, they are adjusted so that each contains a convenient percentage of chi-square values; in other words, so that the probabilities of specific values are easily read. As an example, the last line of column 4 indicates the probability of 0.01 that, in random sampling, chi-square be larger than 6.635.

The distribution of our sample of 230 chi-squares in columns 2 and 3 is readily compared with the theoretical distribution of column 4. Clearly the two do not differ greatly. The discrepancies are due partly to sampling variation, but mainly to this fact: the theoretical distributions in columns 4 and 5 are based on a “continuous” variable, one capable of taking all numerical values, while those in columns 2 and 3 arise from counting a

TABLE 1.4
FREQUENCY DISTRIBUTIONS OF CHI-SQUARE WITH UNEQUAL CLASS INTERVALS

1	2	3	4	5
Class Interval	Discontinuous Variable		Continuous Variable	
	Distribution of 230 Sample Values	Percentage Distribution of Column 2 *	Theoretical Percentage Distribution	Theoretical Cumulative Percentage Distribution
0. -0.000157	0	0	1	100
0.000157-0.000628	0	0	1	99
0.000628-0.00393	11	4.78	3	98
0.00393 -0.0158	20	8.70	5	95
0.0158 -0.0642	10	4.35	10	90
0.0642 -0.148	28	12.17	10	80
0.148 -0.455	47	20.44	20	70
0.455 -1.074	42	18.26	20	50
1.074 -1.642	30	13.04	10	30
1.642 -2.706	22	9.56	10	20
2.706 -3.841	14	6.09	5	10
3.841 -5.412	2	0.87	3	5
5.412 -6.635	1	0.44	1	2
6.635 -	3	1.30	1	1
Total	230	100.00	100	

* Each class frequency of column 2 is expressed as a percentage of 230.

“discontinuous” or “discrete” variable which jumps from one integral number to the next, leaving out intermediate values (fractions and irrationals). You may have observed, for example, that no sample of sizes which you have drawn can yield a chi-square in either of the first two class intervals, but if samples of 500 or more were counted an occasional one would have these small values.

Column 5 contains a *cumulative frequency distribution* of the percentages in column 4. Beginning at the foot of this column 5, each entry is the sum of all the preceding ones in column 4, hence the name. The interpretation is illustrated thus: the fourth-to-last entry, 10%, means that 10% of all samples in the theoretical distribution have chi-squares greater than 2.706. This is the measure of chi-square that we have been needing to enable us to say precisely which values are to be considered small and which large. The discussion will be completed in the next section.

1.13—Sample evidence about the hypothesis. Significant chi-square. It has been emphasized that the function of a sample is to furnish evidence about the population sampled. First we learned that the sample ratio is designed as an unbiased estimate of the ratio in the population. Next we found that the confidence interval contains the population percentage unless a specified caprice of sampling has occurred. Now we have

acquired another piece of evidence associated with some hypothesis which the sampler may have about the population ratio—the probability of getting a random sample more divergent than the one in hand.

Columns 1 and 5 of table 1.4 indicate that 50% of sample chi-squares from a given population have values larger than 0.455. We may, therefore, look upon this as an average value, exceeded as often as not in sampling. On the contrary, chi-squares greater than 3.841 are rare, while those beyond 6.635 occur on the average only once per hundred samples. Did you get values as large as these in your sampling?

It has become customary to refer to any chi-square larger than 3.841 as *significant*. In section 1.10 attention was called to the fact that such large values in your sampling signify no more than unusually divergent samples; but to the investigator, none too sure of his hypothesis, they may suggest that perhaps the sample is not really drawn from the population which he specified. Naturally, if he knew the real ratio in his population, as you do in yours, he would not be doing the experiment. The research worker often designs his samplings for the purpose of testing some hypothesis, accepting or rejecting it on the evidence supplied by the sample. A *significant* value of chi-square denotes a sample so discrepant as to bring into doubt the hypothesis set up; in other words, any chi-square beyond 3.841 is *large* and suggests rejection of the hypothesis. The larger the value of chi-square, the stronger the evidence against the hypothesis.

Before continuing this discussion we must anticipate a bias in associating the theoretical cumulative distribution (table 1.4, column 5) with values of chi-square calculated from enumeration data. This bias leads to an underestimate of the cumulative percentage (or probability). Elimination of the bias is not important in large samples (200 or more) but it is so easily done that one might as well get into the habit of it when testing hypotheses. Merely calculate an *adjusted value* of chi-square, decreasing the absolute value of $X - m$ by 0.5 before squaring, then read the table as usual. For an example, consider a sample of 10 digits with 8 of them odd. Since the expected number of odds is 5.06, the deviations are

$$\begin{aligned} 8 - 5.06 &= 2.94 \\ 2 - 4.94 &= -2.94 \end{aligned}$$

Subtracting 0.5 from each, ignoring the sign, and substituting the results in the formula for chi-square,

$$\text{Adjusted } \chi^2 = \frac{(2.44)^2}{5.06} + \frac{(-2.44)^2}{4.94} = 2.38,$$

instead of the unadjusted value, 3.46. The adjustment is even more easily applied in the formula of section 1.9. It becomes

$$\text{Adjusted } \chi^2 = (|X - m| - 0.5)^2(1/m_1 + 1/m_2),$$

where the symbol, $|X - m|$, indicates the absolute value of the deviation, sign ignored. Substituting the data above,

$$\begin{aligned}\text{Adjusted } \chi^2 &= (2.94 - 0.5)^2 / (1/5.06 + 1/4.94) \\ &= (2.44)^2 (0.40006) = 2.38,\end{aligned}$$

as before.

You will notice that adjusted chi-square is always less than the unadjusted value. This leads to the reading of unbiased probabilities in column 5, probabilities that are greater than would be read from unadjusted chi-squares. Note that the hypothesis is less liable to rejection when this correction of the tabular probability is made. For theoretical discussions of the correction for continuity see references (3), (5), (7), (8), (16).

Look back, now, at the values of chi-square calculated in sections 6, 7, and 8. The one in the election sample, $\chi^2 = 4$ (the adjusted value is 3.96) would be considered unusual if the electorate were really divided equally. It is likely that the sampler, observing this significant value of chi-square, would reject the hypothesis of a tie vote and would say that the election will go to the party which received the majority vote in his sample of 10,000. In saying this he will risk being wrong, because there are almost five chances in a hundred of getting a sample with chi-square greater than 3.96 in sampling from an equally divided population. Such is the kind of risk always run by those who base decisions on sampling.

In the experiment with tomatoes, chi-square was 1.33, the adjusted value being 1.20. Since larger values than this would occur in some 27% of samples drawn from the hypothetical population, the geneticist would probably hold to the 3:1 hypothesis, saying that his chi-square is non-significant. In doing this he may be in error. Actually, the sample ratio was 310:90, and if there were no genetic theory to guide him, he would accept this as his estimate of the population value. Also, he knows well that there are various perturbations of the simple inheritance patterns, and he is on the lookout for them. Nevertheless, in the absence of other information, and in view of the design of the experiment, which was to test the 3:1 hypothesis, he would usually interpret the small chi-square as indicating only sampling variation from this genetic ratio.

Some people adopt a rather slavish attitude toward tests of significance, rejecting the hypothesis if chi-square is more than 3.841 and accepting it if chi-square is less. This indicates inadequate appreciation of the nature of the information acquired by sampling. A sample furnishes evidence, not proof. This evidence is to be added to that already accumulated from experience and reports of other research. Usually, also, there is collateral information accruing during the progress of the experiment. It is the investigator's responsibility to integrate all this evidence and to reach a decision. He cannot evade this responsibility merely by citing a value of chi-square.

EXAMPLE 1.20—Calculate the adjusted value of chi-square for the sample of 800 pea seeds of example 1.15. Ans. 7.80. The geneticist would most likely reject the 1:1 hypothesis for the population from which this sample was drawn.

EXAMPLE 1.21—Calculate adjusted chi-square for the sample of 1,000 voters in example 1.18. Ans. 10.21. The sample constitutes convincing evidence that the candidate's hypothesis was wrong.

EXAMPLE 1.22—In some experiments on heredity in the tomato, MacArthur (10) counted 3,629 fruits with red flesh and 1,176 with yellow. This was in the F_2 generation where the expected ratio was 3:1. Compute $\chi^2 = 0.68$. MacArthur concluded that "The discrepancies between the observed and expected ratios are not significant."

EXAMPLE 1.23—In the same series of experiments, MacArthur counted 671 plants with green foliage and 569 with yellow. This was a backcross in which the theoretical ratio was 1:1. Chi-square = 8.23 is strong evidence against the 1:1 hypothesis, "resulting without doubt from a lower viability in the recessive class."

EXAMPLE 1.24—In the South Dakota farm labor sample (section 1.2) 480 of the 1,000 farmers were classed as owners (or part owners), the remaining 520 being renters. It is known that, of nearly 7,000 farms in the region, 47% are owners. Assuming this to be the population percentage, calculate adjusted chi-square for the sample of 1,000. Ans. $\chi^2 = 0.36$. Does this increase your confidence in the randomness of the sampling? Such collateral evidence is often cited. The assumption is that if the sample is shown to be representative for one attribute it is probably representative also of the attribute under investigation, provided the two are related.

EXAMPLE 1.25—James Snedecor (13) tried the effect of injecting poultry eggs with female sex hormones. In one series 2 normal males were hatched together with 19 chicks which were classified as either normal females or individuals with pronounced female characteristics. What is the probability of the ratio 2:19, or one more extreme, in sampling from a population with equal numbers of the sexes? Ans. $\chi^2 = 12.2$, $P = 0.0004$. For a method of evaluating probabilities not in the table, see section 8.8.

EXAMPLE 1.26—In table 1.4, $42 + 30 + 22 + 14 + 2 + 1 + 3 = 114$ samples have chi-squares greater than 0.455, whereas 50% of 230 were expected. What is the probability of drawing a more discrepant sample if the sampling is truly random? Ans. $\chi^2 = 0.00435$, $P = 0.95$. What is the probability of greater deviation from half-and-half than the deviation in your samples drawn in accord with section 1.5?

The reader who is primarily interested in measurement statistics may wish now to turn forward to chapter 2, omitting the remainder of this one. The necessary preparation has been made in the foregoing sections. He could well return later to the rest of chapter 1 for more thorough study.

1.14—Chi-square and sample size. From the formula,

$$\chi^2 = S(X - m)^2/m,$$

it is clear that if the deviation, $X - m$, is unchanged chi-square decreases as the size of sample increases, because, for any particular hypothesis, m varies directly as sample size. This is illustrated by the first example in table 1.5. Multiplying sample size, n , by four reduces chi-square to one-fourth its former value. The meaning is that, while a deviation of 10 is unusual in a sample of 100, it is rather common in the larger sample—the larger the sample, the more latitude there is for sample variation.

In the second example the deviation is increased fourfold along with sample size, the effect being to multiply the original chi-square by four. The deviation, 40, in the larger sample is much less likely than is 10 in the original. This shows that the latitude for sample variation does not increase directly with sample size.

The deviation in example three is only doubled as the original sample size is quadrupled, chi-square remaining unchanged; that is, the deviation, 20, in a sample of 400 has the same expectation as the deviation, 10, in

the sample one-fourth as large. Thus, for equal likelihood of occurrence the deviation must vary directly as the square root of n .

Clear understanding of the illustrated principles is necessary if one is to appreciate the meaning of chi-square. Here is one application: results are often reported not in numbers originally counted but as percentages of individuals having the attribute; that is, as so many *per hundred* enumerated. It is clear that such percentages cannot be used directly in the calculation of chi-square except in the case where the sample size is an even hundred. In all other samples, before chi-square is computed the percentage must be converted to the actual number of individuals found

TABLE 1.5
THREE EXAMPLES ILLUSTRATING CHANGES IN UNADJUSTED CHI-SQUARE RESULTING FROM QUADRUPLING THE ORIGINAL SAMPLE SIZE

	Original Random Sample	Example 1 Deviation Unchanged	Example 2 Deviation Increased Directly as n	Example 3 Deviation Increased as $\sqrt{n}$
Sample size, n	100	400	400	400
Number having attribute	60	210	240	220
Expected on basis of 1:1 hypothesis	50	200	200	200
Deviation	10	10	40	20
Chi-square (unadjusted)	4	1	16	4

with the attribute, or else some appropriate device must be employed for accomplishing the same end.

Another application of the principles of this section is made in planning the size of a sampling. Consider a circumstance like this: a geneticist studying heredity in the tomato has examined 100 seedlings and found 46 with yellow foliage. Theoretically the plants are expected to segregate half green and half yellow, but he suspects lower viability in the seeds having the recessive yellow gene. The deviation of yellow seedlings from the expected 50 is 4% of the number examined, but chi-square (unadjusted) is only 0.64, so the statistical evidence against the 1:1 ratio is slight. Nevertheless, the suspicion of lower viability must be tested, and the question is asked, "In the event that the 4% deviation is found in a larger sample, how many plants should be examined to make chi-square 10 times its present sample value?" The appropriate example in table 1.5 is the second because, the deviation being assumed to be 4% of the plants counted, its size is to increase directly as n . If so, chi-square will increase in the same way, and a sample of 10 times 100 should have the required unadjusted value, 6.4. MacArthur's data in example 1.23 might have been the outcome of some such reasoning, verifying the suspected character of the gene. On the other hand, the suspicion might have proved un-

founded, the larger sample turning up with no more than ordinary deviation from the 1:1 ratio. In either event the investigator has acquired the desired information.

1.15—Another formula for chi-square. This formula gives the same result as the more general one, but often is easier of computation:

$$\chi^2 = \frac{(a - rb)^2}{r(a + b)},$$

where a and b are the numbers in the two classes and r is the hypothetical ratio of the corresponding numbers in the population. Applying this to the tomato problem of section 1.6:

$$\begin{aligned} a &= \text{number of red fleshed fruits} = 310 \\ b &= \text{number of yellow fleshed fruits} = 90 \\ a + b &= 400 \\ r &= 3/1 = 3 \end{aligned}$$

$$\chi^2 = \frac{(310 - 3 \times 90)^2}{3(400)} = 1.33, \text{ as before.}$$

The adjustment of chi-square is introduced rather simply:

$$\begin{aligned} \chi^2 &= \frac{(|a - rb| - (r + 1)/2)^2}{r(a + b)} \\ &= \frac{(310 - 270 - (3 + 1)/2)^2}{3(400)} \\ &= 1.20, \text{ as in section 1.13} \end{aligned}$$

Returning to the question of sample size, the formula of this section furnishes a convenient answer if the investigator, wishing to test the $r:1$ hypothesis, is willing to rest his decision on a sample whose deviation is any amount up to $p\%$ of the sample size. This size is to be chosen so that if the deviation turns out to be $p\%$ then chi-square shall be significant. His thinking would be that any less deviation is inconsequential, but that if the $p\%$ deviation proves significant the fact will be worth knowing. With a little manipulation, the formula becomes

$$n = \frac{r\chi^2}{[(1 + r)p/100]^2},$$

where chi-square may be chosen at any desired level. As an example, suppose the 3:1 hypothesis is to be tested, and a deviation of 5% of n will be interesting if significant. Then $r = 3$, $\chi^2 = 3.841$ and $p = 5\%$. Substituting,

$$n = \frac{3(3.841)}{(4 \times 0.05)^2} = 288$$

If 288 plants are examined, and if the deviation from the expected 3:1 segregation is 5% of 288 = 15 plants or more, then chi-square will be adjudged significant.

EXAMPLE 1.27—Jenkins and Bell (9) reported a gene producing yellow seedlings in maize, a simple recessive to the normal green. Among 9,717 counted the ratio of green to yellow seedlings was 78.95% to 21.05%, whereas the expected ratio was 3:1. Does the sample evidence support the hypothetical 3:1 ratio? Ans. Adjusted $\chi^2 = 80.83$. The investigators showed by means of further experiments that the gene was "linked with lethal or semi-lethal factors which caused a deficiency in the yellow seedling class. Satisfactory 3:1 ratios were not obtained until these cultures had been outcrossed and the disturbing factors eliminated."

1.16—Ratios, rates, and percentages. These fractions are very popular in the presentation of statistics because, being usually reduced to some common denominator, their numerators are readily compared. Examples are given in table 1.6. The percentages of males all have the denominator, 100 guinea pigs. The two sets of ratios are numerators, respectively, of *one* and *one hundred* females. Other denominators are sometimes used. Death rates, for example, are usually expressed as so many per thousand people, or per 100,000 for some rare diseases.

Ease of comparison in a set of these fractions is gained at the expense

TABLE 1.6
AVERAGE NUMBERS OF MALES AND FEMALES LITTERED IN A COLONY OF GUINEA PIGS,
BY MONTH. FROM SCHOTT AND LAMBERT (12)
Various Sex Ratios

	Jan.	Feb.	Mar.	Apr.	May	June
Males	65	64	65	41	72	80
Females	49	58	81	48	62	80
Total	114	122	146	89	134	160
Percentage males	57.0	52.5	44.5	46.1	53.7	50.0
Ratio, males to females	1.327	1.103	0.802	0.854	1.161	1.000
Ratio, males per 100 females	132.7	110.3	80.2	85.4	116.1	100.0

	July	Aug.	Sept.	Oct.	Nov.	Dec.	Total
Males	88	114	80	129	112	86	996
Females	95	118	94	104	144	85	1,018
Total	183	232	174	233	256	171	2,014
Percentage males	48.1	49.1	46.0	55.4	43.8	50.3	49.45
Ratio, males to females	0.926	0.966	0.851	1.240	0.778	1.012	0.9784
Ratio, males per 100 females	92.6	96.6	85.1	124.0	77.8	101.2	97.84

of information. The original denominators are set aside, even sometimes lost sight of, a fault to be deplored. Many computations require return to the original data, as for example the fractions in the column of totals where:

$$49.45\% = (100)(996)/2,014, \text{ etc.}$$

Attention has already been called to the fact (section 1.14) that percentages cannot be used in the calculation of chi-square unless some method is provided for recovery of the original base, thus giving due weight to sample size.

In many instances the bases of comparable percentages are not very different so that the percentages themselves can be averaged with no great inaccuracy. Even the guinea pig percentages of males have an average,

$$\frac{57.0 + 52.5 + \dots + 50.3}{12} = \frac{596.5}{12} = 49.71\%$$

which is not greatly divergent from the correct 49.45%. However, if one wishes an accurate average of percentages that have different bases he must go back to the original fractions, add their numerators and denominators separately, then divide.

EXAMPLE 1.28—Three dairy herds of a certain community showed the following reaction to a test for tuberculosis:

Number cows in herd	40	100	10
Percentage reactors	5	2	60

Calculate the average, 6.7%. Do you think this is a better average than $(5 + 2 + 60)/3 = 22.3\%$?

EXAMPLE 1.29—The percentages of noxious weed seeds in two samples of timothy are 0.01% and 0.05%. If each sample consisted of 10,000 seeds, what is the average percentage in the two? Ans. 0.03%.

EXAMPLE 1.30—If the samples in the foregoing example were 80,000 and 20,000 seeds, respectively, what would be the average? Ans. 0.018%, quite properly nearer the percentage of the larger sample.

EXAMPLE 1.31—Schott and Lambert reported that the numbers in table 1.6 are averages over 7 years so that the total number of males was 6,972 and of females, 7,126, the sex ratio being 97.84 males per 100 females. Test the hypothesis that the population sex ratio is 100. Ans. $\chi^2 = 1.66$. **NOTE:** If the averages were used, chi-square would be 0.219, only about one-seventh the correct value. Why isn't it exactly one-seventh?

[1.17—Suggestions and cautions. Summary. In reporting your sampling or experiment, bear in mind these principles. (i) The estimates, confidence statements and tests of hypotheses presumably summarize all the information in your data; consequently there is seldom need for publication of the numerical details. (ii) Space in journals is precious and should not be used for describing computational procedures or explaining the meaning of statistical terms. Those are functions of the

textbook. Your job is to report the facts of the sampling, to discuss the relation of your findings to the problem set up, to make decisions and draw conclusions. (iii) A well-constructed table often displays all the desired information about the statistics of an investigation. It leaves both the writer and reader free to consider the biological implications, unhampered by numerical detail or such distracting comment as "the difference is statistically significant." (iv) The sampled population is specified by a description of the circumstances surrounding the experiment (compare section 1.2). Details of the species and varieties used, the breeding and rearing of animals, soil, fertilizer and season for plants, laboratory techniques—such are required in order that other data from the same population may be compared to yours with no more than allowable sampling differences between them.

There is widespread misinformation concerning sample based inferences about populations. It may be well to point out two misstatements often met. First, about the 0.95 confidence interval, or more particularly about the similar concepts of standard and probable error, one often hears it said that "95% of other samples from the same population will have ratios lying within this interval." If you look carefully through sections 1.3 and 1.4, especially through your sampling of section 1.5, you will find no basis for such an inference. Only if the population ratio is known, as it was in the calculations of section 1.9, can the distribution of sample ratios be predicted by the methods which have been presented in this chapter. The fallacy of the statement is evident from your sampling of digits. For illustration, select one of the samples with a percentage rather divergent from 50.6%, then calculate the proportion of your other sample percentages falling within the confidence interval set. Perhaps less than half of them do. You may find some samples with percentages close to 50% for which the statement is true, but these are exceptions.

Another common but erroneous inference follows from the chi-square test. If chi-square is small it is often said that "the odds are 19:1 that the hypothesis is correct." I have presented no method for evaluating the probability of hypotheses. From the viewpoint adopted in studying chi-square the hypothesis is fixed, the probability evaluated being that of drawing a sample from it more extreme than the one in hand. It is true that after the evidence is all accumulated the investigator must make a decision about the hypothesis, and his decision has some presumably high probability of being correct, but we do not know how to evaluate such a probability.

Summarizing this chapter, we have studied the sample as a source of information about a population. If randomly drawn it provides an unbiased estimate, a confidence interval and a test of the sampler's hypothesis. By drawing a lot of samples we have verified some useful mathematical theorems about them, notably, the confidence interval and the distribution of chi-square. The chi-square coefficient of dispersion has

proved itself a powerful tool for extracting information about the population.

In chapters 9 and 16 we shall return to the consideration of enumeration data. In those intervening, methods appropriate to measurements will occupy attention.

REFERENCES

1. HERBERT ARKIN and RAYMOND R. COLTON. *Graphs: How to Make and Use Them*. Harper and Brothers, New York and London (1936).
2. C. J. CLOPPER and E. S. PEARSON. *Biometrika* 26:404 (1934).
3. W. G. COCHRAN. *Annals of Eugenics*, 7:207 (1936).
4. R. A. FISHER. *Journal of the Royal Statistical Society*, 85:87 (1922).
5. R. A. FISHER. *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh (1941).
6. R. A. FISHER and F. YATES. *Statistical Tables*. Oliver and Boyd, Edinburgh (1938).
7. H. C. FRYER and JOHN W. GOWEN. *Genetics*, 27:212 (1942).
8. C. H. GOULDEN. *Methods of Statistical Analysis*. John Wiley and Sons, Inc., New York (1939).
9. MERLE F. JENKINS and MARTIN A. BELL. *Genetics*, 15:253 (1930).
10. JOHN W. MACARTHUR. *Transactions of the Royal Canadian Institute*, 18:1 (1931).
11. KARL PEARSON. *Philosophical Magazine*, Series 5, 50:157 (1899).
12. R. G. SCHOTT and W. V. LAMBERT. *Iowa State College Journal of Science*, 4:343 (1930).
13. JAMES SNEDECOR. Personal communication, (1943).
14. RAY E. WAKELEY. *Iowa State College Research Bulletin* 249 (1938).
15. HELEN M. WALKER and N. DUROST. *Statistical Tables; their structure and use*. Teachers College, Columbia University, New York (1936).
16. F. YATES. *Supplement to the Journal of the Royal Statistical Society*, 1:217 (1934).
17. G. U. YULE. *Journal of the Royal Statistical Society*, 85:95 (1922).

Chapter 2

MEASUREMENT DATA. COMPARISON OF INDIVIDUALS

2.1—Introduction. In the first chapter only one kind of data was considered, that resulting from the enumeration of individuals possessed of some interesting attribute. Alive or dead, infested or free, red or white—such were the categories observed. Another type of data, even more common, derives from measuring some continuous variate like height or yield or income. How much do you weigh? How old are you? Numerical facts like these are intimately related to all our thinking. In this and following chapters will be developed statistical methods appropriate to such measurement data. The principle involved is the same as that already presented: the collection and examination of a sample for the purpose of learning about the population. Unbiased estimates, confidence statements and tests of hypotheses will be the familiar concepts employed.

2.2—The mean as an unbiased estimate. Let us turn immediately to some reported data and look later at the pertinent theory. In 1936 the Council on Foods of the American Medical Association sampled the vitamin C content of commercially canned tomato juice by analyzing a specimen from each of the brands that stood accepted by the Council (4). The data are shown in the second column of table 2.1. A commonly used average, the *arithmetic mean* or simply *the mean*, is calculated by dividing the sum of the 17 items by their number,

$$333/17 = 19.6 \text{ milligrams of ascorbic acid per 100 grams of juice}$$

This average summarizes the results in a single value which, in some respects, is typical of the whole set. In size it is intermediate among the individual data. In successive samples from the same population the mean varies less than the observations, one of the reasons that give us confidence in using it. Another of these reasons is that, with the assumptions to be described below, this sample average is an unbiased estimate of the population mean. We have thus reached again the familiar first inference of section 1.3—the mean vitamin C content of the tomato juice manufactured in 1936 under the 17 specified brands is estimated without bias by the sample mean, 19.6 mg./100 g. To make the second inference (section 2.12) some further preparation is necessary.

2.3—The range, a measure of variation. In looking through the vitamin C contents of the seventeen brands, one is struck by their variability. From high to low, their *range* is

$$29 - 13 = 16 \text{ mg. per 100 grams}$$

TABLE 2.1
VITAMIN C CONTENT OF 17 SPECIMENS OF COMMERCIALY CANNED
TOMATO JUICE (4), 1936

Observation Number	Vitamin C Content Mg. per 100 g.		Deviation From Mean	Deviation Squared
n	Original	X	$x = X - \bar{x}$	x^2
1	16	16	- 4	16
2	22	22	+ 2	4
3	21	21	+ 1	1
4	20	20	0	0
5	23	23	+ 3	9
6	22	21	+ 1	1
7	17	19	- 1	1
8	15	15	- 5	25
9	13	13	- 7	49
10	22	23	+ 3	9
11	17	17	- 3	9
12	18	20	0	0
13	29	29	+ 9	81
14	17	18	- 2	4
15	22	22	+ 2	4
16	16	16	- 4	16
17	23	25	+ 5	25
Totals	333	340	-26 +26	254

$\bar{x} = 340/17 = 20$ mg. per 100 grams
 $s^2 = 254/16 = 15.88$ $s = 3.98$ mg./100 g.
 $s_x^2 = 15.88/17 = 0.934$ $s_x = 0.97$ mg./100 g.

What are the causes of this variation? Different processes of manufacture, perhaps, and different sources of the fruit. Doubtless, also, the specimens examined, being themselves samples of their brands, differed from the brand means. Finally, the laboratory technique of evaluation is never perfectly accurate. Variation is the very essence of statistical data, and the range of a sample is a good rough-and-ready way to measure it. A more reliable estimate will be given in section 2.8 below.

The mean and the range constitute a summary of the information contained in the entire set. The average vitamin C content is 19.6 mg. per 100 grams with a range of 16 mg. per 100 grams. These two summary numbers convey the outstanding facts of the sample.

2.4—The array. To make succeeding computations simpler let us temporarily transfer attention to the slightly modified series in the third column of the table. Some of the more intimate features of this sample may be revealed by arranging it from low to high in an *array* like this:

13, 15, 16, 16, 17, 18, 19, 20, 20, 21, 21, 22, 22, 23, 23, 25, 29

The range is now obvious. Attention is also attracted to the *concentration* of the values near the center of the array. It is this *central tendency* of sets of data that gives rise to that faith we feel in samples and in estimates based on them.

2.5—Graphical representation. Instead of presenting data in a table, it is sometimes more effective to display the facts graphically. Figure 2.1 is a graphical representation of the foregoing array of 17 vitamin determinations. A dot represents each item. The distance of the dot from the vertical line at the left, proportional to the concentration

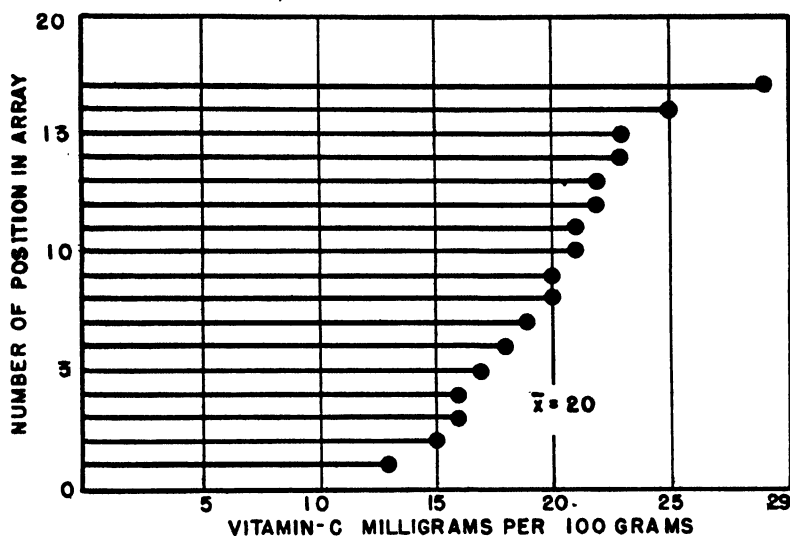


FIG. 2.1—Graphical representation of an array. Modified Vitamin C data.

of ascorbic acid in a brand specimen, is read in milligrams per 100 grams on the horizontal scale.

The diagram brings out vividly not only the variation and the concentration in the sample, but also two other characteristics: (i) the rather symmetrical occurrence of the values above and below the mean, and (ii) the scarcity of both extremely small and extremely large vitamin C contents, the bulk of the items being near the middle of the set. These features recur with notable persistence in measurement data as well as in the samples of enumeration data examined in chapter 1. For many variables associated with living organisms there are averages and ranges peculiar to each, reflecting the manner in which each seems to express itself most successfully. These norms persist despite the fact that individuals enjoy a considerable freedom in development. A large part of our thinking is built up around ideas corresponding to such statistics. The words, pig, daisy, man, each raises an image which is quantitatively described by summary numbers. It is difficult to conceive of any progress in thought until memories of individuals are collected into concepts like averages and ranges.

2.6—Symbolical representation. The items in any set may be represented by

$$X_1, X_2, X_3, \dots X_n,$$

where the subscripts 1, 2, $\dots$ n , may specify position in the set of n items. (The three dots accompanying these symbols are read, *and so on*.) Matching the symbols with the values in column 3 of table 2.1,

$$X_1 = 16, X_2 = 22, \dots X_{17} = 23 \text{ mg./100 g.}$$

The mean is represented by $\bar{x}$, so that

$$\bar{x} = (X_1 + X_2 + \dots + X_n)/n$$

This is condensed into the form

$$\bar{x} = (SX)/n,$$

where X stands for every item successively. The symbol, SX , is read, "summation X " or "sum of the X 's" while $\bar{x}$ is called "x-bar," or "bar-x." Applying this formula to the values of X in table 2.1,

$$SX = 340, \text{ and } \bar{x} = 340/17 = 20 \text{ mg./100 g.}$$

The population mean may be symbolized by m in order to distinguish it from the sample estimate, $\bar{x}$ (compare the use of m in section 1.8).

2.7—Deviations from mean. The individual variations of the items in a set of data may be well expressed by the *deviations* of these items from some centrally located number such as the mean. For example, the deviation-from-mean of the first X -value in table 2.1 is $16-20 = -4$ mg. per 100 g.; that is, this specimen falls short of the mean by 4 mg./100 g. Of special interest is the whole set of deviations calculated from the array in section 2.4:

$$-7, -5, -4, -4, -3, -2, -1, 0, 0, 1, 1, 2, 2, 3, 3, 5, 9$$

These deviations are represented graphically in figure 2.1 by the distances of the dots from the vertical line drawn through the mean.

Deviations are almost as fundamental in our thinking as are averages themselves. "What a whale of a pig" is a metaphor expressing astonishment at the deviation of an individual's size from the speaker's concept of the normal. Gossip and news are concerned chiefly with deviations from accepted standards of behavior. Curiously, interest is wont to center in departures from norm, rather than in that background of averages against which the departures achieve prominence. Statistically, freaks are freaks only because of their large deviations.

Deviations are represented symbolically by lower case letters. That is:

$$\begin{aligned} x_1 &= X_1 - \bar{x} \\ x_2 &= X_2 - \bar{x} \\ &\vdots \\ x_n &= X_n - \bar{x} \end{aligned}$$

Just as X may represent any of the items in a set, or all of them in succession, so x represents deviations from mean. In general,

$$x = X - \bar{x}$$

It is easy to verify the theorem that the sum of a set of deviations-from-mean is zero; that is $\sum x = 0$. The set listed in table 2.1 adds to zero, the sum of the positive deviations being equal to the sum of the negatives. This theorem about deviations-from-mean is useful for verifying the calculation of a set of deviations—be sure that the sum is zero. As a consequence of the theorem, it follows that the *mean* of the deviations is zero. This is a theorem about which you will be reminded later.

EXAMPLE 2.1—The weights of 12 staminate hemp plants in early April at College Station, Texas, (17), were approximately:

13, 11, 16, 5, 3, 18, 9, 9, 8, 6, 27, and 7 grams

Array the weights and represent them graphically. Calculate the mean, 11 grams, and the deviations therefrom. Verify the fact that $\sum x = 0$.

EXAMPLE 2.2—The heights of 11 men are 64, 70, 65, 69, 68, 67, 68, 67, 66, 72, and 61 inches. Compute the mean and verify it by summing the deviations. Are the numbers of positive and negative deviations equal, or only their sums?

EXAMPLE 2.3—The yields of alfalfa from 10 plots were 0.8, 1.3, 1.5, 1.7, 1.8, 2.0, 2.0, 2.0, 2.0, and 2.2 tons per acre. How many deviations are positive and how many negative? Is their sum zero?

✓ **EXAMPLE 2.4**—The weights of 11 forty-year-old men were 148, 154, 158, 160, 161, 162, 166, 170, 182, 195, and 236 pounds. Contrast the graphical representation of this array with that of the preceding example. Notice the fact that only three of the weights exceed the mean. Why? Is the mean typical of this sample of weights? Is $\sum x = 0$?

EXAMPLE 2.5—The following were the yields of two varieties of oats in five successive years (bushels per acre):

Variety	Year				
	1	2	3	4	5
A	34	30	41	25	45
B	30	15	33	25	25

Compute the mean difference, 9.4 bushels per acre. Show that the sum of the deviations of the differences from the mean difference is zero.

EXAMPLE 2.6—The following data are adapted from Reddy's (14) investigations of the differences in yield attributable to the disinfection of *Diplodia* infected seeds of maize. The figures represent bushels per acre.

Treatment	Pairs of Plots in 1933					
	1	2	3	4	5	6
Treated	68.1	74.6	64.4	69.2	61.8	57.9
Untreated	64.7	62.5	66.8	69.2	53.9	58.5
Differences	3.4	12.1	-2.4	0.0	7.9	-0.6

Pairs of Plots in 1934									
1	2	3	4	5	6	7	8	9	10
18.0	24.0	18.8	17.8	18.5	27.2	23.6	23.9	20.3	11.9
10.9	24.4	15.1	16.8	13.2	21.6	13.7	17.5	16.3	15.5
7.1	-0.4	3.7	1.0	5.3	5.6	9.9	6.4	4.0	-3.6

Verify the fact that the mean difference is (approximately) 3.7 bushels per acre. Represent graphically the array of differences. It is worthy of note that the differences are rather uniformly distributed throughout the two years despite the great disparity in the seasons' yields.

✓**EXAMPLE 2.7**—If you sum the deviations from 3.7 bushels per acre in the foregoing example you will not get zero. Why? If you compute the mean of the deviations and add it to 3.7, will you get the exact mean of the differences? Are you keeping clearly in mind the distinction between deviations and other differences?

EXAMPLE 2.8—Suppose you wish to estimate the yield of a field of 300 rows of corn. You actually harvest 10 rows and determine the mean yield, 5 bushels per row. Would you hesitate to fix the field yield at 1,500 bushels? You would be assuming that the field mean is the same as that of the 10 harvested rows, and would be using the theorem, $\sum X = n\bar{x}$.

EXAMPLE 2.9—If you have some skill in algebra, prove the theorem that $\sum x = 0$. Starting with the relation, $x = X - \bar{x}$, sum both members, then substitute $\sum X = n\bar{x}$.

EXAMPLE 2.10—If you have two sets of data which are paired as in example 2.5, and if you have calculated the resulting set of differences, prove that the mean of the differences is equal to the difference between the means of the two sets. Verify this theorem by use of the data in example 2.5.

2.8—The standard deviation. We have mentioned the range as a measure of variation; a good one, too, if no very discriminating work is to be done. A moment's thought will convince you, however, that it is not reliable. Just look back at figure 2.1 and call to mind the infrequency of occurrence of extreme measurements. There are not many of them in the whole population, making it improbable that they will be well represented in a small sample. Many samples, in fact, will include none of the population oddities, being made up entirely of those more numerous individuals nearer the mean. You see, the range, depending wholly upon the difference between the two sample extremes, is likely to be erratic in sampling. For exacting investigations it is desirable to have a more stable average of variation, one resembling the mean in characteristics and method of computation.

With this ideal in mind, one naturally turns to the set of deviations in section 2.7. Surely, an average of variation should be founded on these measures of individual variation. Since their mean, however, is always zero, that simple average is worthless. A natural suggestion is to ignore the signs, computing the mean of the *absolute values* of the deviations. This average, the *mean absolute deviation*, had a considerable vogue in times past, but it was soon found to lead into a blind alley so far as statistical theory is concerned.

After much experience, statisticians selected an average of variation called *standard deviation*, symbolized by s . Its calculation is set out in the right hand part of table 2.1. First, each deviation is squared. Next, the *sum of squares*, Sx^2 , is divided by the number of *degrees of freedom*, $(n - 1)$, yielding the *mean square* or *variance*, s^2 . Finally, the extraction of the square root recovers the original unit of measurement, milligrams per 100 grams of juice, so the standard deviation is a typical deviation in much the same way that the mean is a typical observation.

2.9—Why square the deviations? In a non-mathematical discussion, it is quite impossible to give an adequate answer to this question. The most convincing justification of the process must lie in the value of the results produced, and these will become increasingly evident as your knowledge of statistics expands. Nevertheless, some excuses may be in order. Attention was called to the facts that neither the deviations nor their absolute values furnished satisfactory averages. Squaring the deviations is the next operation in respect of simplicity. Incidentally, the squares are all positive, increasing the ease of computation. Some people think that squaring makes harder calculations than absolute values would. While this is true in the simple form of table 2.1, it is exactly contrary to the facts in the more extensive calculations that will be required later. The squares lend themselves amazingly to short cuts in computation and powerful methods in statistics.

More fundamentally, squares of deviations are basic in the theory of the best known and most widely used frequency distribution, the *normal*. You will hear more of this distribution in later chapters. It is symmetrical, contrasting sharply with that of chi-square (section 1.11). The values in table 2.1 and the differences in example 2.6 may well be samples drawn from normally distributed populations. It is the widespread incidence of this and related populations in biology that accounts for the well-nigh universal use of squared deviations for making the average of variation.

People with a vivid sense of form find help in understanding the nature of standard deviation by considering the geometric interpretation of the computations of table 2.1. Draw to scale 17 straight lines, one representing each of the deviations. The zero deviations may be symbolized by points. Construct a square on each of the lines as side. The areas of these squares correspond to the squares of the deviations in the last column of the table. The sum of all the areas is 254 square units. The average area, the mean square, is 15.88 square units. The standard deviation, the square root of 15.88, is the side of a square having the average area, and is measured in the same unit as are the 17 deviations and the 17 observed values.

2.10—Why divide by $(n - 1)$? This is another question whose answer lies in the mathematical theory, but which has implications important to the biologist. Two of these will be mentioned here. The

first has to do with the sample itself, consisting of n independent observations. From these the mean, $\bar{x}$ (a *statistic*), is computed and set up as an estimate of the mean, m (a *parameter*), of the population. Then deviations are calculated. These, however, are not all independent. Since their sum must be zero, after $(n - 1)$ of them are selected at random the n th is fixed. Suppose, for example, four out of five deviations are taken quite arbitrarily as 3, -1 , 2, and 5. The fifth, then must have the value,

$$0 - (3 - 1 + 2 + 5) = -9$$

This is an explanation of the inference made above that there are only $(n - 1)$ *degrees of freedom* associated with the sum of squares, Sx^2 , the n th degree of freedom being assigned to the average from which deviations are measured. The mean square, or variance, is therefore defined as $Sx^2/(n - 1)$.

The second implication is equally important. If we were concerned only with a description of sample variance, we might still insist on dividing the sum of squares by n . But our prime interest is in the population variance. How can we obtain an unbiased estimate of that parameter? The answer is to divide the sum of squares by the number of degrees of freedom, $(n - 1)$, instead of by the sample size, n . One explanation is easy to find. It lies in the improbability of drawing representatives of the population extremes in small samples. The range of the sample is likely to be much less than that of the population. The sample mean square deviation (variance) would likely fall short of the parameter if division were by n . "Student" (16) showed in 1908 that this tendency is compensated, on the average, if the denominator $(n - 1)$ is used. While the distinction is numerically unimportant in large samples, it is vital in much experimentation.

EXAMPLE 2.11—Verify the facts that the mean of the following set is 10 grams and the standard deviation, 4 grams:

3, 6, 6, 12, 9, 12, 10, 9, 12, 14, 17

EXAMPLE 2.12—Compute the standard deviation, 6.7 grams, of the weights of hemp plants in example 2.1. Note that somewhat more than two-thirds of the weights fall within the interval 11.0 ± 6.7 grams.

2.11—Some characteristics of the standard deviation. The squaring of the deviations has two slightly inconvenient results: the distinction between positive and negative deviations is lost, and the few large deviations contribute a disproportionate amount to the sum of squares. It is imperative to understand clearly the consequent characteristics of the standard deviation. Like any other average it is, in a specific way, typical of the deviations it represents, and is intermediate among them in size. Let us look at these deviations (section 2.7), shorn of their signs and rearranged:

0, 0, 1, 1, 1, 2, 2, 2, 3, 3, 3, 4, 4, 5, 5, 7, 9 mg./100 g.

There is no symmetry about this set. It might well have been drawn from a population distributed like chi-square. Certainly, the mean of these deviations (that is, the mean absolute deviation) will not be centrally located in the set. Both these results follow the loss of the signs. Furthermore, the two largest absolute deviations contribute more than half of the total sum of squares (table 2.1). The consequence is that 11 of the 17 absolute deviations are smaller than s (3.98 mg./100 g.) and only 6 larger.

Much the same account can be given without the bother of absolute deviations. Look back at the array of X in section 2.4. Consider the interval from

$$\bar{x} - s = 20 - 3.98 = 16.02 \text{ mg./100 g.}$$

to

$$\bar{x} + s = 20 + 3.98 = 23.98 \text{ mg./100 g.}$$

This interval includes the 11 values,

17, 18, 19, 20, 20, 21, 21, 22, 22, 23, 23, mg./100 g.

These 11 items constitute 65% of the total of 17. Such a majority of observed values within the interval, $\bar{x} \pm s$, is common in sampling. It reflects the fact that in normally distributed populations with mean, m , and standard deviation, σ (the Greek lower case *sigma* is used to distinguish the population parameter from the sample estimate, s), the interval, $m \pm \sigma$, contains 68.27% of the items; that is, about two-thirds. Of course, it may be expected that in random samples, even those drawn from normal populations, this fraction may vary considerably. Nevertheless, one should always take the precaution to verify his computation of s by observing the proportion of items falling between $\bar{x} - s$ and $\bar{x} + s$. If it is radically different from two-thirds, mistakes in arithmetic are suggested.

If you study mathematical statistics, you may hear a good deal about the Principle of Least Squares. The mean and deviations therefrom are related to that principle in this manner: if deviations are measured from the mean, then the sum of their squares is a minimum. In particular (and reversing the statement) if deviations in table 2.1 are measured from some number different from the mean, 20, the sum of their squares will be greater than 254. Verify this by trying deviations from, say, 19 then 22.

This least squares theorem about the mean throws further light on division by $n - 1$. Since the mean of a sample is usually different from that of the population, the sum of squares of deviations from the sample mean is commonly less than the corresponding sum of squares of deviations from the population mean. In sampling from a normal distribution, division by $n - 1$ counterbalances this tendency, giving an unbiased estimate of the population variance. The standard deviation, s , is not an unbiased estimate of the *population standard deviation*, σ , but tends to be smaller. This is one reason for using the variance so much, another being its greater ease of computation.

As would be expected, the range and standard deviation are not unrelated. If you draw a lot of samples from a normally distributed population, each sample with 9 or 10 items, you will find that the average of their ranges is about three times the population standard deviation. Hence, one-third the range of each such sample is an estimate of σ in the population. Similarly, if n is between 25 and 30, one-fourth the range is the appropriate estimate of σ (see table 5.5). By fixing in mind a few such values (section 5.8), one has a basis for estimating the standard deviations of familiar populations. Even more, he can make fairly good guesses at the standard deviations, s , of the samples observed. As an example, suppose a group of 27 young men of age 22 are randomly chosen insofar as stature is concerned. Measure the heights of the shortest and tallest of the group and subtract to get the range, say

$$73 - 62 = 11 \text{ inches.}$$

Not only is one-fourth of this range, 2.8 inches, an estimate of the standard deviation of the population of 22-year-old men, known to be about 2.7 inches, but it is also a rough estimate of the standard deviation of the sample. This process is valuable for two reasons: it serves as a check on computations of s , and it enables one to translate his common knowledge of range into the new quantity, standard deviation.

It seems rather characteristic that large things vary much and small things little. For this reason it is often convenient to express the standard deviation as a fraction of the mean, the resulting statistic being called *relative standard deviation* or *coefficient of variation*, C . As an example, it is reported (1) that the average statures of one-year and eighteen-year girls are 74.4 and 161.0 cm. respectively, with standard deviations 2.64 and 6.12 cm. The two coefficients of variation are

$$\begin{aligned} C_1 &= 2.64/74.4 = 0.036 \\ C_{18} &= 6.12/161.0 = 0.038, \end{aligned}$$

almost the same. Usually C is expressed as a percentage, C_1 for example, being 3.6%. Discussion of this characteristic of the standard deviation is resumed in section 2.15.

EXAMPLE 2.13—The birth weights of 20 guinea pigs, borne in litters of two, were:

30, 30, 26, 32, 30, 23, 29, 31, 36, 30, 25, 34, 32, 29, 28, 27, 38, 31, 34, 30 grams.

In computing the deviations subtract, not the exact mean, but the nearest integer, 30 grams. This will save squaring a lot of decimals. Is the sum of deviations zero? What is the mean of the deviations? How does your standard deviation compare with the correct value, 3.60 grams? What proportion of the birth weights lie in the interval 30.25 ± 3.60 grams?

EXAMPLE 2.14—The standard deviation of the differences in yield, example 2.6, is 4.46 bushels per acre. For computation, carry only one decimal place in the mean, then compare your result with that just stated. For ordinary purposes it is quite all right to retain in the mean only the precision of the observations. More precise results can be got with little extra labor by the methods of chapter 5.

EXAMPLE 2.15—Would there be any object to calculating the mean and standard deviation of the 16 yields from treated seed in example 2.6? Would they be estimates of the parameters of any population? Suppose 10 plots had been harvested in 1933 and 6 in 1934, would the estimates be altered? Would the populations be affected?

EXAMPLE 2.16—In samples of 15 to 17 items the range is, on the average, about 3.5 times the standard deviation. Verify this statement from the sample of assays in table 2.1. How does it compare with the facts in example 2.6? The statement may be written in any of these three ways:

$$\begin{aligned}\text{range} &= 3.5s \\ (\text{range})/s &= 3.5 \\ (\text{range})/3.5 &= s\end{aligned}$$

For samples containing numbers of items different from 15 to 17, see section 5.7.

EXAMPLE 2.17—The mean yield of hay from 15 plots of alfalfa was 2.2 tons per acre, with $s = 0.35$ ton per acre. Would you estimate the range of the 15 plot yields as 1.2 tons per acre? Would this suggest that the highest yielding plot bore about 2.8 tons per acre?

EXAMPLE 2.18—Suppose you lined up according to height the 16 men in two squads of 18-year-old freshmen, then measured the height of the shortest, 63 inches, and of the tallest, 72 inches. Would you expect the standard deviation of the 16 heights to be about 2.6 inches? Would you accept the midpoint of the range, 67.5 inches, as a rough average height?

EXAMPLE 2.19—In a normally distributed array of the gains in weight of 15 swine during a 120-day feeding experiment, the mean gain was 180 pounds. Two-thirds of the gains fell within the interval from 145 to 215 pounds. Do you agree that the standard deviation was probably about 35 pounds? If so, would you estimate the least of the 15 gains at about 120 pounds?

EXAMPLE 2.20—If you like puzzles, try making an approximately normal set of 16 items with $\bar{x} = 20$ and $s = 4$. It is easier to start with a set of deviations. Test your result by counting the number of items in the interval 20 ± 4 , and by comparing the range with $3.5s$.

2.12—The confidence or fiducial inference. After occupying the foregoing nine sections with concepts necessary to making this second inference about the population, we shall still have to anticipate some ideas that will be developed in the next chapter. This is in marked contrast to the sequence in chapter 1, where the second inference followed immediately after the first (section 1.3). Looking forward, then, we borrow a formula and a table (section 3.6 and table 3.8) which will enable us to set limits to the accuracy of the mean, 20 mg./100 g., of the sample of vitamin C determinations. The formula,

$$s_{\bar{x}} = s/\sqrt{n},$$

enables us to estimate the *standard deviation of means* of samples drawn from some single population. Means are expected to vary less than items. The formula indicates this and gives the quantitative relation: the *standard error*, as $s_{\bar{x}}$ is usually called, is inversely proportional to the square root of the sample size, n . For the vitamin C,

$$s_{\bar{x}} = 3.98/\sqrt{17} = 0.97 \text{ mg./100 g.,}$$

as recorded at the end of table 2.1. This says that if many samples of 17 were drawn from the same population, and if the standard deviation of

their means were calculated as in table 2.1, the answer is estimated from the single sample as 0.97 mg./100 g. It is to be noted that means of samples of 17 vary about one-fourth as much as the items themselves.

Turning now to table 3.8, the quantity t is a ratio defined by the formula,

$$\pm t = (\bar{x} - m)/s_{\bar{x}}$$

We know $\bar{x}$ and $s_{\bar{x}}$ from our sample and can select an appropriate value of t from the table; therefore, we can calculate

$$m = \bar{x} \pm ts_{\bar{x}}$$

Now, the sampling distribution of t , like that of χ^2 , has been worked out so that one may select a value corresponding to many convenient probabilities. If we choose the 5% level, for example, and substitute it along with $\bar{x}$ and $s_{\bar{x}}$ in the equation above, there will result 2 values of m which are called *fiducial limits*. Designating them as l_1 and l_2 , we have:

$$l_1 = \bar{x} - t_{.05}s_{\bar{x}} \quad \text{and} \quad l_2 = \bar{x} + t_{.05}s_{\bar{x}}$$

Finally, if we say that the population mean, m , lies between these limits, the probability of being misled by the sampling is only 0.05.

In the vitamin C sample, $\bar{x} = 20$ and $s_{\bar{x}} = 0.97$ mg./100 g.; for 16 degrees of freedom, $t_{.05} = 2.120$; hence,

$$\begin{aligned} l_1 &= 20 - (2.120)(0.97) = 18 \text{ mg./100 g., and} \\ l_2 &= 20 + (2.120)(0.97) = 22 \text{ mg./100 g.} \end{aligned}$$

We may assert, then, that in the sampled population the average vitamin C content is between 18 and 22 mg./100 g., and we shall be right unless a one-in-twenty mischance has occurred in the sampling.

Fisher (7) designates the interval from l_1 to l_2 as the 0.95 *fiducial interval*. This corresponds to the 0.95 confidence interval of table 1.1, and the inference now made about the population mean is of the same general nature as that learned in chapter 1.

What are the theoretical conditions that must be met if the first two inferences are to be correct? Random sampling from a single, normally distributed population is one way to state them (in technical language, these are sufficient but not necessary conditions). It is hard to substitute any more general statement instead of the requirement of randomness, so let's stick to that. But normal distribution of the population is a specification that can be considerably relaxed. This is fortunate, because we seldom know the distribution of sampled populations. From time to time in what follows, attention will be called to devices appropriate to samples from non-normal distributions. Otherwise, we shall assume that abnormality affects our inferences but little.

Are the theoretical conditions realized in the vitamin C sampling? The published report is not in sufficient detail to warrant judgment. Each manufacturer sent a specimen for analysis, but no information is given as

to the manner of selection. We assume random choice. As for normality, there is certainly no evidence against it in the sample (for tests of normality, see chapter 8). Finally, the requirement of sampling from a single population was probably not included in the plans. The vitamin *C* content of each brand may have had its own population mean and standard deviation, different from those of the others, and the quantities of juice marketed under the several brands probably differed. The design and statistical analysis of samples from such groups of populations will be discussed in chapter 17. At present we can say only this: our inferences about the population are exact if the theoretical conditions were met. As has been mentioned before, it is likely that all practical samplings depart more or less from the mathematical models. If the investigator plans his experiment with due regard to theory, then executes it skillfully, his inferences should be reliable approximations.

EXAMPLE 2.21—In examples 2.1 and 2.12 you calculated $\bar{x} = 11$ grams and $s = 6.7$ grams for the April weights of 12 staminate hemp plants. Assuming random sampling from a normal population of weights, compute the 95% fiducial interval as $6.7 - 15.3$ grams. I hope you are impressed by the unreliability of this small sample with a standard deviation more than half the size of the mean ($C = 61\%$).

EXAMPLE 2.22—In an investigation of growth in school children of 8 private schools (8), the mean height of 265 boys of age 13.5 to 14.5 years was 63.84 inches with standard deviation, 3.08 inches. What is the 95% fiducial interval of the mean of the population represented? Ans. 63.5 – 64.2 inches. Calculate $C = 4.8\%$.

EXAMPLE 2.23—Set 99% fiducial limits to the population mean of the foregoing example. Ans. 63.4 – 64.3 inches.

EXAMPLE 2.24—The mean and standard deviation of the original assays in column 2 of table 2.1 are respectively 19.6 and 3.95 mg./100 g. What is the 95% fiducial interval? Ans. 17.6 – 21.6 mg./100 g. Compare this with the interval, 18 – 22 mg./100 g., computed from the modified set in column 3 of the table.

2.13—The comparison of individuals. Test of a null hypothesis. As pointed out in chapter 1, the investigator usually has definite questions about how his experimental material will behave under specific circumstances. To get evidence, he sets up an experiment, more or less complicated. One of the simplest is planned to contrast the behaviors of two similar individuals subjected to different conditions. The individuals to be compared may be field plots or pigs or colonies of bees. But if there were only a single pair it would be impossible to say whether the difference in behavior is to be attributed to the conditions imposed or to the natural variability of the individuals or partly to each. Hence, there must be two or more pairs, or *replications*, one member of every pair being chosen at random to receive the first treatment, the other member the second. The differences between the measurements of the two members of the pairs constitute the sample data upon which inferences are to be based. If there were no individual variation the differences would presumably be all alike. Actually, the variation among the differences is called *experimental error*. The outcome of the experiment is evaluated by comparing the average difference, attributable to treatment, with the *standard error* of this

mean difference. Let us examine such an experiment together with the statistical method for reducing the data.

Youden and Beale (19) wished to measure the difference in concentration of two preparations of the mosaic virus. The method employed was to rub half a leaf of a tobacco plant with a piece of cheesecloth soaked in one preparation of the virus extract, and to rub the other half leaf similarly with the second preparation. The measurement of concentration is the number of local lesions appearing on the half leaf. The data reported in table 2.2 are measurements from leaf No. 2 on each of 8 plants. It is

TABLE 2.2

NUMBERS OF LESIONS ON HALVES OF 8 TOBACCO LEAVES, EACH BEING LEAF NO. 2 ON 8 PLANTS. THE LESIONS MEASURE THE CONCENTRATIONS OF TWO PREPARATIONS OF THE MOSIAC VIRUS *

Plant No.	Lesions on Half Leaves		Difference $X = X_1 - X_2$	Deviation $x = X - \bar{x}$	Squared Deviation x^2
	Preparation 1 X_1	Preparation 2 X_2			
1	9	10	- 1	-5	25
2	17	11	6	2	4
3	31	18	13	9	81
4	18	14	4	0	0
5	7	6	1	-3	9
6	8	7	1	-3	9
7	20	17	3	-1	1
8	10	5	5	1	1
Total	120	88	32	0	130
Mean	15	11	$\bar{x} = 4$ lesions		$s^2 = 18.57$

$$s = \sqrt{18.57} = 4.31 \text{ lesions. } s_2^2 = 18.57/8 = 2.32. s_2 = 1.52 \text{ lesions.}$$

* Slightly changed to make calculation easier.

known that not only do individual tobacco plants react differently to the virus, but also leaves in different positions. Hence, each preparation is applied to every leaf, the halves tending to respond alike if treated alike. Thus, the difference in concentration of the two preparations is measured by the difference between the numbers of lesions as recorded in column 4 of the table. These 8 independent measurements of the difference in concentration constitute a random sample from a population of such differences, and can be summarized in exactly the same way as the sample of vitamin determinations in table 2.1. The mean difference, 4 lesions, is an estimate of the population mean, with $s_2 = 1.52$ lesions as the corresponding standard error.

Remembering one object of the experiment, to learn whether there is a difference in concentration between the two preparations, we now set up the hypothesis that the population mean difference is zero, then ask, "What is the probability of a sample mean larger than 4 lesions in random

samples of 8 differences from the specified population?" To get this information the quantity, t , of table 3.8 is again invoked. Now, by hypothesis, $m = 0$, so

$$t = (\bar{x} - 0)/s_z = 4/1.52 = 2.63$$

This sample value of t is to be compared with the tabular values opposite 7 degrees of freedom. Since 2.63 lies about one-fourth the way from the 5% level to the 1%, the conclusion is that so large a value of t would occur in only 4 per 100 samples from the hypothetical population. As we learned in chapter 1, the null hypothesis is likely rejected, and the two preparations considered different in concentration. You see, the third inference about a population, described rather elaborately in section 1.6 to 1.13, has now been adapted to measurement data. The sampling justification of the process, analogous to that of sections 1.9 to 1.12, is reserved for chapter 3. With the experience acquired in the first chapter, you have probably had no difficulty in following the condensed presentation in this section.

Having concluded that there is a population difference, you may wish to set bounds to it. We have, for $d.f. = 7$, $t_{.05} = 2.365$, and hence

$$l_1 = 4 - (2.365)(1.52) = 0.41 \text{ lesions, and}$$

$$l_2 = 4 + (2.365)(1.52) = 7.59 \text{ lesions}$$

If we say, then, that the difference in population concentration is measured by a number of lesions between 0.41 and 7.59, we shall be right unless a one-in-twenty chance has come off.

EXAMPLE 2.25—Using the statistics of examples 2.6 and 2.14 ($\bar{x} = 3.71$ and $s = 4.46$ bushels per acre, with $n = 16$) calculate $t = 3.33$ under the hypothesis that $m = 0$. Compare this with the 1% level, $t = 2.947$. What conclusion do you draw?

EXAMPLE 2.26—L. C. Grove (9) determined the mean numbers of florets produced by seven pairs of plots of Excellence gladiolus, one plot of each pair planted with high (first year) corms, the other with low (second year or older) corms. The plot means were as follows:

Corm	Florets						
High	11.2	13.3	12.8	13.7	12.2	11.9	12.1
Low	14.6	12.6	15.0	15.6	12.7	12.0	13.1

Compute the differences, the mean difference, 1.2 florets, the standard deviation, 1.4 florets, and t , 2.3, assuming $m = 0$. What is a reasonable inference about the population mean difference?

EXAMPLE 2.27—Mitchell, Burroughs and Beadles (11) computed the biological values of proteins from raw peanuts (P) and roasted peanuts (R) as determined in an experiment with 10 pairs of rats. The pairs of data P, R are as follows: 61, 55; 60, 54; 56, 47; 63, 59; 56, 51; 63, 61; 59, 57; 56, 54; 44, 62; 61, 58. Compute the mean difference 2.1, and the standard deviation of the differences 7.41 units. Since $t = 0.90$, approximately 40 out of 100 similar samples from a population with $m = 0$ would be expected to have larger t -values.

NOTE: 9 of the 10 differences, $P - R$, are positive. One would like some information about the next-to-the-last pair 44, 62. The first member, especially, seems abnormal.

While unusual individuals occur in the most carefully conducted trials, their appearance demands extra attention. You may have the curiosity to recalculate the statistics omitting this divergent pair. It always astonishes me that a single pair can reverse the conclusions.

EXAMPLE 2.28—Crampton (5) reports the gains in weight per 100 pounds of feed eaten for 10 pairs of swine constituting lots I and III of MacDonald College hog feeding trial 33B. The pigs were paired on the basis of being litter mates of as nearly the same initial weight as feasible. The pairs of gains are as follows, the datum for the member of lot I in each pair being set down first: 17, 20; 22, 21; 22, 21; 15, 22; 24, 24; 22, 22; 21, 23; 21, 21; 17, 22; 21, 23 pounds gain per 100 pounds feed. Compute the mean difference, 1.7 pounds, in favor of lot III, and the standard deviation 2.7 pounds. Retain one decimal place in your deviation from mean if you wish to verify this standard deviation. Calculate t , and observe the probability of occurrence of a larger if $m = 0$. Ans. $t = 2.0$ and $P = 0.08$.

EXAMPLE 2.29—Prof. H. H. Love, of Cornell University, took a leading part in introducing "Student's" small sample theory in America. In one of his illustrations (10) he compared the yields of Great Northern (G) and Big Four (B) oats in the nine years 1912–1920. The differences, $G - B$, were 16.3, 13.4, 3.8, 7.9, 2.6, 2.5, 9.6, 7.2, and 3.3 bushels per acre. Compute $\bar{x} = 7.4$, and $s = 5.0$ bushels per acre. The difference in the yields of the two kinds of oats is highly significant.

EXAMPLE 2.30—Salmon (15) mentions a case in which 648 paired cooperative yield tests of Kanred and Turkey winter wheats were conducted in Kansas. It was desirable to make a rough test of significance without computing $\bar{x}$ and s . Instead, the number of tests in which Kanred outyielded Turkey was counted, 429, with six ties. Dividing the ties equally, it may be said that Kanred outyielded Turkey in the ratio of 432:216 tests. This ratio can be tested for significance against the theoretical 1:1, using the methods of chapter 1. The value, $\chi^2 = 71$, makes significance indubitable. Brandt (2) gives a graphical method for making the test.

The reader may well omit the remainder of this chapter on the first reading. He may turn to chapter 3, or if he wishes to acquaint himself first with methods only, he may skip to chapter 4.

2.14—Sample size. By substituting $m = 0$ and $s_{\bar{x}} = s/\sqrt{n}$, the formula for t may be written

$$t = \bar{x}\sqrt{n}/s,$$

from which,

$$n = t^2 s^2 / \bar{x}^2 \quad \text{or} \quad n = t^2 C^2,$$

where C is the coefficient of variation defined in section 2.11. Thus for any specified probability of t , sample size may be estimated if any reasonable assumptions about s and $\bar{x}$ are available.

As an illustration, refer to example 2.27 where statistics for the biological value of proteins from raw peanuts were $\bar{x} = 2.1$, $s = 7.4$, and consequently $C = 7.4/2.1 = 3.5$ (that is, 350%). If one assumes that approximately this same ratio will eventuate in a larger sample, he may determine the number of pairs of rats necessary to produce $t = 2$,

$$n = 2^2(3.5)^2 = 49$$

If this larger sample is drawn, it may turn out to have a value of C either greater or less than 3.5, with t correspondingly either beyond the 1% level or less than the 5%. One should be prepared to reject the null hypothesis or to accept it according to the evidence of the new sample. If the decision is important enough, the substituted value of t^2 may be

increased to 7, corresponding roughly to the 1% level. However, it seems a good guess that the unusual pair, 44, 62, would seldom appear again, and that even 10 pairs would ordinarily indicate a real population difference.

A moment's thought will make it clear that sampling might be continued until an unusually large value of t is turned up even if m is really zero. If one quits when the results tally with his wishes, he modifies the probabilities set down in table 3.8. He might even be misled by a second sample designed as above, because this larger sample might be the freakish one.

It may be well to emphasize again the idea that statistical evidence is not proof. Even after extensive sampling the investigator may not reject the null hypothesis when in fact the hypothesis is false. For example, m may not be zero in the population, yet natural variation may be great enough to confine t to a moderate value in any practicable size of sample. Therefore, when one fails to reject the hypothesis he does not thereby conclude that m is zero. He decides only that m is so small as to be unimportant to his investigation.

2.15—The relative variation. Coefficient of variation. In section 2.11, C was defined as $s/\bar{x}$. As was indicated, its utility lies mainly in the fact that in many series the mean and standard deviation tend to change together. This is illustrated by the mean stature and corresponding standard deviation of girls from 1 to 18 years of age shown graphically in figure 2.2. Until the twelfth year the standard deviation increases at somewhat greater rate, relative to its mean, than does stature, causing the coefficient of variation to rise, but by the seventeenth year and thereafter C is back to where it started. Without serious discrepancy one may fix in mind the figure, $C = 3.75\%$, as the relative standard deviation of human stature, male as well as female. More precisely, the coefficient rises rather steadily from infancy through puberty, falls sharply during a brief period of uniformity, then takes on its permanent value near 3.75%.

A knowledge of relative variation is valuable in planning and in evaluating experiments. In the preceding section was an example of the utility of C in estimating sample size. After the statistics of an experiment are summarized, one may judge of its success partly by looking at C . In corn variety trials, for example, although mean yield and standard deviation vary with location and season, yet the coefficient of variation is often between 5% and 15%. Values outside of this interval cause the investigator to wonder if an error has been made in calculation, or if some unusual circumstances throw doubt on the validity of the experiment. Similarly, each sampler knows what values of C may be expected in his own data, and is suspicious of any great deviation.

Other uses of the coefficient of variation are numerous but less prevalent. Since C is the ratio of two averages having the same unit of measurement it is itself independent of the unit employed. Thus, it is the same

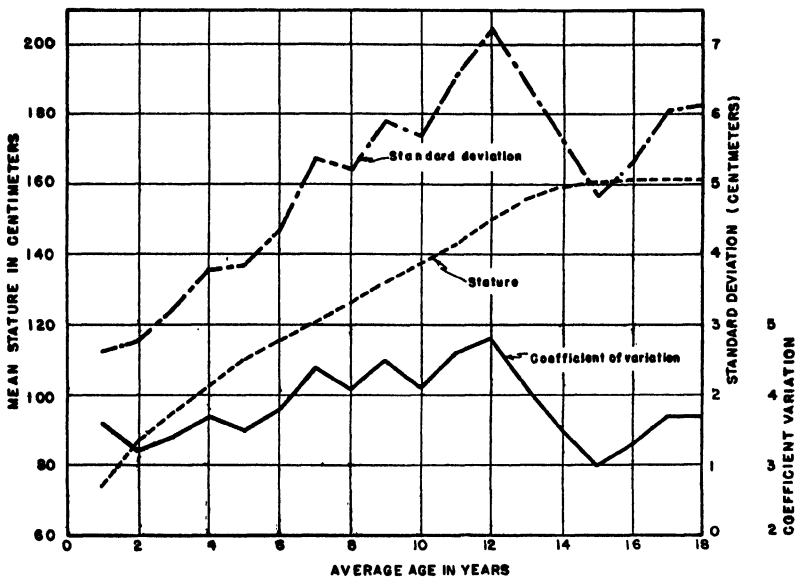


FIG. 2.2—Graph of 3 time series; stature, standard deviation and coefficient of variation of girls from 1 to 18 years of age. See reference (1).

whether inches, feet or centimeters are used to measure height. Also, the coefficient of variation of the yield of hay is comparable to that of the yield of corn.

Experimental animals have characteristic coefficients of variation, and these may be compared despite the diversity of the variables measured. The body weight of male albino rats between 90 and 243 days of age has an average C of about 14% (6), while that for annual egg production of Barred Plymouth Rock hens is close to 32% (13). Such information is essential to those planning experiments.

Like many other ratios, the coefficient of variation is so convenient to use that some people overlook the information contained in the original data. Try to imagine how limited you would be in interpreting the stature-of-girls coefficients if they were not accompanied by $\bar{x}$ and s . You would not know whether an increase in C is due to a rising s or a falling $\bar{x}$, nor whether the saw-tooth appearance of the C -curve results from irregularities in one or both of the others, unless indeed you could supply the facts from your own fund of knowledge. The coefficient is informative and useful in the presence of $\bar{x}$ and s , but abstracted from them it may be misleading.

EXAMPLE 2.31—In example 2.28 the mean difference between the gains of ten pairs of pigs was 1.7 pounds with standard deviation, 2.7 pounds. Assuming the same

coefficient of variation in a following experiment, how many pairs should be included to bring t to its 5% level? Ans. Approximately 13 pairs.

EXAMPLE 2.32—In experiments involving chlorophyll determinations in pineapple plants (18), the question was raised as to the method that would give the most consistent results. Three bases of measurement were tried, each involving 12-leaf samples, with the statistics reported below. From the coefficients of variation, it was decided that the methods were equally reliable, and that the most convenient one could be chosen with no sacrifice of precision.

STATISTICS OF CHLOROPHYLL DETERMINATIONS ON 12-LEAF SAMPLES FROM PINEAPPLE PLANTS, USING THREE BASES OF MEASUREMENT

Statistic	100-gram Wet Basis	100-gram Dry Basis	100-sq. cm. Basis
Mean (milligrams)	61.4	337	13.71
Standard deviation (milligrams)	5.22	31.2	1.20
Coefficient of variation (percent)	8.5	9.3	8.8

EXAMPLE 2.33—In well-conducted experiments with properly developed and adequately fed growing swine, the standard deviation of gains in weight may be taken as 10% of the mean gain. If you are told that a lot of pigs gained 150 pounds during a certain period, would you expect the standard deviation of the gains to be about 15 pounds? Would it be equally true that an average daily gain of 1.5 pounds per day should be accompanied by a standard deviation of approximately 0.15 pound per day?

EXAMPLE 2.34—In a certain laboratory there is a colony of rats in which the coefficient of variation of the weights of males between 56 and 84 days of age is close to 13%. Estimate the standard deviation of the weights of a lot of these rats whose mean weight is 200 grams. Ans. 26 grams.

EXAMPLE 2.35—Assuming that the coefficient of variation of yields in field plot tests with wheat is usually near 5%, would you be surprised if told that in an experiment where the yield was 25 bushels per acre, the standard deviation of plot yields was 0.5 bushel per acre?

EXAMPLE 2.36—In example 2.17, the coefficient of variation of the plot yields is 16%. Do you know enough about alfalfa to judge whether that is a reasonable figure?

2.16—Probable deviation. This is an average of variation, widely used in the past but now rapidly being replaced by the standard deviation. In size, it is approximately two-thirds of the standard deviation. As for computation, one calculates the standard deviation first, as in table 2.1, then multiplies by 0.6745 to get the probable deviation. The reason for the factor, 0.6745, (or approximately, $\frac{2}{3}$) will be set out in chapter 8.

What are the advantages of the probable deviation, $p.d.$, over the standard? Only one. In a normally distributed population, the interval, $m \pm p.d.$, contains half the items. It is perhaps easier to think of one-half than of the 0.6827 associated with the interval $m \pm \sigma$ (section 2.11). This advantage is so far overbalanced by the many superiorities of the standard deviation as to make it seem likely that the probable deviation will disappear from use in biological statistics.

In random samples drawn even from normally distributed populations

the proportion of items within the interval, $\bar{x} \pm p.d.$, varies considerably. Examine the heights of men in example 2.2. Since $s = 3$ inches the probable deviation is $0.6745 \times 3 = 2.02$ inches. The interval, 67 ± 2.02 inches extends from 64.98 to 69.02 inches. Within this interval lie the 7 heights, 65, 66, 67, 67, 68, 68, 69, constituting 64% of the entire sample number. Contrariwise, only seven of the sixteen differences in example 2.6 lie in the interval, 3.7 ± 3.0 bushels per acre (see example 2.14). However, in larger samples drawn from normally distributed populations, one often finds approximately 50% of the items in the specified interval.

It is necessary to become familiar with the probable deviation in order to read technical articles involving statistics. Remember, probable deviation is only about $\frac{3}{8}$ of standard; while standard deviation, the larger, is 1.5 times the probable.

2.17—The median. The mean is the average of concentration peculiarly appropriate for describing samples from normally distributed populations. When populations depart notably from the normal, the mean may not be at all typical. In such cases, another average of concentration, the *median*, may be used for description, supplementing the mean. The median is the middle item in an array.

As an example of its utility in description, look back at the weights of 11 forty-year-old men listed in example 2.4. Contrast the graph of this array, figure 2.3, with that of the vitamin data of figure 2.1. Note that only three of the weights exceed the mean, 172 pounds. The median, 162 pounds, is the sixth weight in the array, five weights being larger and five smaller. In a sense, the median describes the set of weights in better fashion than does the mean. Together, they give a fairly adequate account of the central tendency of the weights in this sample. The difference, (median) — (mean), is sometimes used as a rough measure of the skewness of a sample. The disability of the mean, alone, as an average of skewed sets is that it is unduly influenced by the few extreme items.

Weights of mature individuals are usually distributed in asymmetrical fashion, the majority of them lying near the lower end of the array. Apparently, individuals find it difficult to survive with weights much lower than normal, but may carry far greater weights without immediate penalty. Skewed distributions result from any limit of variation at one end of the range. For this reason, dressing percentages of swine are apt to be concentrated in the upper part of the array. Asymmetrical distributions constitute a difficult problem in statistics. We shall avoid them as much as possible until we reach a later chapter.

If the median is used as the average of concentration, it is customary to describe the variation of a set by fixing two other values called *quartiles*. Just as the median is the item at the middle of an array, so the quartiles lie at the middle of the lower and upper halves of the array. In the array of men's weights, the third weight, 158 pounds, is the first quartile, Q_1 ,

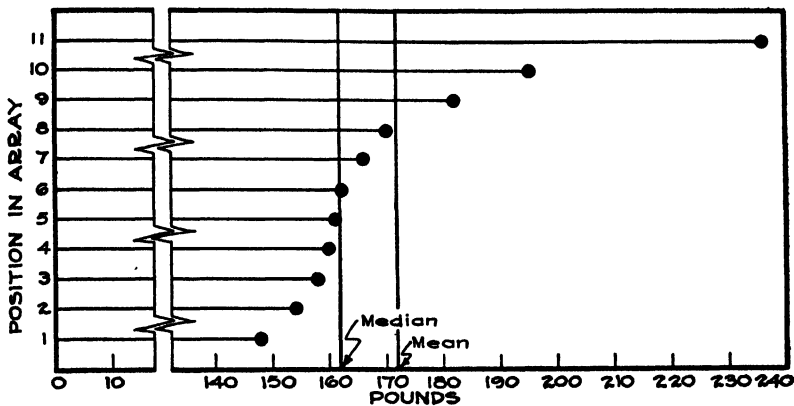


FIG. 2.3—Weights of 11 forty-year-old men. The break attracts attention to the fact that the differences in weight are exaggerated in order to emphasize the asymmetry of the array.

while the ninth, 182 pounds, is the third quartile, Q_3 . The median, then, might be called the second quartile, Q_2 . The *inter-quartile* interval, $Q_3 - Q_1$, is a measure of variation. Like the interval, $\bar{x} \pm p.d.$ in the normal distribution, the interval $Q_3 - Q_1$ in every distribution contains 50% of the items. In a normally distributed sample, therefore, $Q_3 - Q_1 = 2p.d.$

In a sample like that of example 2.39, there are no items exactly dividing the set into fourths. The median is then defined as the interpolated value of item number $(n + 1)/2$, which, in this case, is item number $(24 + 1)/2 = 12\frac{1}{2}$. The interpolated value is the mean of the two items on either side of this one, the twelfth and thirteenth. Its value is, then, $(19.9 + 20.1)/2 = 20$, the same as the mean. Similarly, Q_1 is the interpolated value of item number $(n + 1)/4 = 6\frac{1}{4}$. This value divides the interval between the sixth and seventh items in the ratio, 1:3. Q_1 is therefore a *weighted mean* of these items, $[3(15.3) + 16.3]/4 = 15.55$. For Q_3 , the item number is $3(n + 1)/4 = 18\frac{3}{4}$, and the interpolated value,

$$[23.7 + 3(24.7)]/4 = 24.45$$

The inter-quartile interval is $24.45 - 15.55 = 8.9$, which is about as near to $2p.d. = 8.4$ as one ordinarily gets in these approximately normal samples.

The median-quartile description of a sample falls into the same category as the range, furnishing a rough-and-ready means of summarizing the data. So far as more advanced statistical methods are concerned, it is not often useful.

EXAMPLE 2.37—Verify the fact that the interval, $\bar{x} \pm p.d.$, contains about half the values of X in table 2.1.

EXAMPLE 2.38—The mean height of 840 thirteen-year-old boys is reported as 57.3 inches with probable deviation 1.86 inches. Would you expect to find about 420 of these heights between 55.4 and 59.2 inches? Compute $s = 2.76$ inches and $C = 4.8\%$.

EXAMPLE 2.39—The following approximately normal sample has the mean, 20.0: 7.7, 9.9, 11.6, 13.0, 14.2, 15.3, 16.3, 17.2, 18.0, 18.8, 19.5, 19.9, 20.1, 20.5, 21.2, 22.0, 22.8, 23.7, 24.7, 25.8, 27.0, 28.4, 30.1, 32.3. Estimate the probable deviation by measuring the interval, equally spaced on either side of the mean, within which half the items lie. Would you begin measuring from 16.3 or from 15.8? Using the method of table 2.1, $s = 6.3$, and thence, probable deviation = 4.2. Do two-thirds of the items lie within the interval, 20 ± 6.3 ?

EXAMPLE 2.40—Try writing a normal sample of 16 items with mean 50 and probable deviation 10. Don't forget example 2.16. Establish your success by using the method of computation in table 2.1.

EXAMPLE 2.41—Show that, of the differences in example 2.6, $Q_1 = -0.30$, $Q_3 = 3.85$, and $Q_2 = 6.92$ bushels per acre.

EXAMPLE 2.42—In the sample of 11 weights of example 2.4 and section 2.16, would you expect to find $Q_3 - Q_1 = 2p.d.$? It is easy to test your expectations.

EXAMPLE 2.43—The weights of neutral fats in the blood plasma of 64 normal men (12) were 419, 162, 149, 219, 248, 313, 211, 169, 91, 281, 264, 172, 124, 235, 94, 62, 224, 58, 92, 205, 132, 145, 305, 285, 174, 107, 240, 269, 396, 416, 662, 703, 249, 179, 136, 157, 198, 95, 100, 178, 145, 199, 54, 407, 166, 94, 248, 235, 66, 120, 239, 128, 560, 233, 80, 557, 217, 542, 252, 175, 103, 165, 351, 107 mg. per 100 cc. Array these weights and represent them graphically. Calculate the mean, 224.4 mg. per 100 cc. The median is 188.5 mg. per cc. Can you give an explanation of the skewness of the distribution?

EXAMPLE 2.44—Calculate the deviations from the median for the 11 weights of 40-year-old men. Is the sum equal to zero? Is the sum of the deviations from the mean equal to zero?

EXAMPLE 2.45—If you ignore the signs of the deviations in the preceding example, using only their absolute values, which is the larger, the sum of the absolute deviations from mean or median?

EXAMPLE 2.46—Clements and Long (3), in discussing the effects of light on the growth of sunflower plants, said that the difference in stature produced by lights in various intensities ranged from 14.8 to 32.0 cm., with mean equal to 23.3 cm. The *midpoint* of the range, $(14.8 + 32.0)/2 = 23.4$ cm., is almost the same as the mean. Do you think this is convincing evidence of symmetry in the heights of the sunflower plants?

EXAMPLE 2.47—Observe that in figure 2.3 the midpoint of the range is 192 pounds greater than either median or mean. Why?

2.18—Description of a sample. Summary. The entire information in a sample from a normally distributed population is summarized by the three facts: number of observations, mean, and standard deviation. Little is gained by reporting the individual items in either table or graph. A statement of the fiducial interval may be helpful to the reader, as well as the probability associated with whatever hypothesis has been set up. One advantage of these summary numbers is that all the relevant information from the sample is conveyed briefly, leaving time and space for discussion of the biological implications.

In this text, attention has been focused chiefly on inferences about the population from which the sample was taken. The mean and variance of a random sample from a normal population as unbiased estimates of the

corresponding parameters, the fiducial interval as an indication of the reliability of the sampling, and the test of some appropriate hypothesis—these are the inductions which can be made from sample to population.

The replicated comparison of individuals is a simple and effective experimental design for testing the difference between two treatments. The two members of a pair would presumably behave alike if treated alike, and the difference between their behavior measures the differential treatment effect. Replication tends to average out variation if the individuals of each pair are randomly assigned to the treatments. The variation among the pair differences is a measure of the experimental error inherent in such a trial.

The natural extension to replicated *sets* of individuals, each assigned to one of three or more treatments, will be made in chapter 11. Meanwhile, we must pause to develop the theory which we have already found essential for both fiducial statements and tests of hypotheses.

REFERENCES

1. BERNICE BOYNTON. University of Iowa Studies in Child Welfare, Vol. 12, No. 4 (1936).
2. A. E. BRANDT. Journal of the American Statistical Association, 28:434 (1933).
3. FREDERIC E. CLEMENTS and FRANCES C. LONG. Plant Physiology, 9:779 (1934).
4. COUNCIL ON FOODS. Journal of the American Medical Association, 110:651 (1938).
5. EARL WILCOX CRAMPTON. Journal of Nutrition, 7:305 (1934).
6. HENRY H. DONALDSON. The Rat. The Wistar Institute of Anatomy and Biology, No. 6. Philadelphia (1924).
7. R. A. FISHER. Proceedings of the Royal Society of London, A, 139:343 (1933). Proceedings of the Cambridge Philosophical Society, 26:528 (1930).
8. HORACE GRAY and J. C. AYRES. Growth in Private School Children. The University of Chicago Press, Chicago (1931).
9. L. C. GROVE. Iowa Agricultural Experiment Station Bulletin 253 (1939).
10. H. H. LOVE. Journal of the American Society of Agronomy, 16:64 (1924).
11. H. H. MITCHELL, WISE BURROUGHS and JESSE R. BEADLES. Journal of Nutrition, 11:257 (1936).
12. IRVINE H. PAGE, ESBEN KIRK, WILLIAM H. LEWIS, JR., WILLIAM THOMPSON and DONALD D. VAN SLYKE. The Journal of Biological Chemistry, 111:613 (1935).
13. RAYMOND PEARL and F. M. SURFACE, USDA, Bureau of Animal Industry Bulletin, No. 110, Part I (1909).
14. CHAS. S. REDDY. Iowa State College Journal of Science, 9:530 (1935).
15. S. C. SALMON. The Journal of the American Society of Agronomy, 22:77 (1930).
16. "STUDENT." Biometrika, 6:1 (1908).
17. PAUL J. TALLEY. Plant Physiology, 9:737 (1934).
18. R. K. TAM and O. C. MAGISTAD. Plant Physiology, 10:161 (1935).
19. W. J. YOUDEN and HELEN PURDY BEALE. Contributions from Boyce Thompson Institute, 6:437 (1934).

Chapter 3

SAMPLING FROM A NORMALLY DISTRIBUTED POPULATION

3.1—Introduction. In chapter 1 the facts about confidence intervals were verified through sampling and this same device outlined the chi-square distribution which led to desired probabilities of occurrence. But in chapter 2 you were asked to accept s_x^2 and t with almost no justification. Now we are ready to examine their distributions along with those of $\bar{x}$, s^2 , and s , in the expectation of learning the reasons for the statements already made about them.

There are two devices for studying the sampling properties of statistics. The first is that employed by the mathematician who arrives at the results through use of his symbolical logic. He furnishes the formulas and tables that make statistical methods accurate and convenient. The second device is the actual drawing of samples. This requires no mathematical training but makes it possible to verify many findings of probability theory. More important, the layman is enabled by his experience in sampling to appreciate the foundations upon which the mathematical logic is built.

Both the mathematician and the experimenter have been eager to draw samples from actual populations which depart more or less from the mathematical models. The data may come from harvesting a crop in small plots, with perhaps a thousand measurements of some such character as yield (1) (5) (6). The *geographical distribution* of the variable is then known as well as the *frequency distribution*. The latter may be compared with the normal distribution, often assumed to be the source of samples, while the former is the basis for studying the effectiveness of various experimental designs.

For our purposes, we shall, as before, set up a particular kind of distribution and draw samples from it with the help of the table of random digits. But now, instead of having only two kinds of individuals to deal with, as in the enumeration statistics which have been studied, we are concerned with a population whose members theoretically have all values of some variate like weight. In this model distribution there is no limit to the number of different weights, but practically the scale for measuring can be read no closer than a certain *least count*. Hence, the sample that we have devised to simulate a normal population departs from the model in two respects; it is limited in size and range instead of being infinite, and has a *discontinuous* variate in lieu of the continuous one implied in the theory. The effects of these departures will scarcely be noticed because they are small in comparison with sampling variation.

3.2—A normally distributed sample. In table 3.1 are the weight gains of a hundred swine, slightly modified from experimental data so as to form a distribution which is approximately normal. The items are numbered from 00 to 99 in order that they may be identified easily with corresponding numbers taken from the table of random digits. The salient features of this kind of distribution may be discerned in figure 3.1. The

TABLE 3.1
ARRAY OF GAINS IN WEIGHT (POUNDS) OF 100 SWINE DURING A PERIOD OF 20 DAYS
The gains approximate a normal distribution with
 $m = 30$ pounds and $\sigma = 10$ pounds

Item Number	Gain	Item Number	Gain	Item Number	Gain	Item Number	Gain
00	3	25	24	50	30	75	37
01	7	26	24	51	30	76	37
02	11	27	24	52	30	77	38
03	12	28	25	53	30	78	38
04	13	29	25	54	30	79	39
05	14	30	25	55	31	80	39
06	15	31	26	56	31	81	39
07	16	32	26	57	31	82	40
08	17	33	26	58	31	83	40
09	17	34	26	59	32	84	41
10	18	35	27	60	32	85	41
11	18	36	27	61	33	86	41
12	18	37	27	62	33	87	42
13	19	38	28	63	33	88	42
14	19	39	28	64	33	89	42
15	19	40	28	65	33	90	43
16	20	41	29	66	34	91	43
17	20	42	29	67	34	92	44
18	21	43	29	68	34	93	45
19	21	44	29	69	35	94	46
20	21	45	30	70	35	95	47
21	22	46	30	71	35	96	48
22	22	47	30	72	36	97	49
23	23	48	30	73	36	98	53
24	23	49	30	74	36	99	57

gains, clustering at the midpoint of the array, thin out symmetrically, slowly at first, then more and more rapidly. Two-thirds of the gains lie in the interval 30 ± 10 pounds, that is, in an interval of two standard deviations centered on the mean. In a real population, indefinitely great in number of individuals, greater extremes doubtless would exist, but that need cause us little concern.

The relation of the histogram to the array is clear. After the class bounds are decided upon, it is necessary merely to count the dots lying between the vertical lines, then make the height of the rectangle proportional to their number. The central value, or *class mark*, of each interval is indicated on the horizontal scale of gains.

In table 3.2 is the frequency distribution graphically represented in

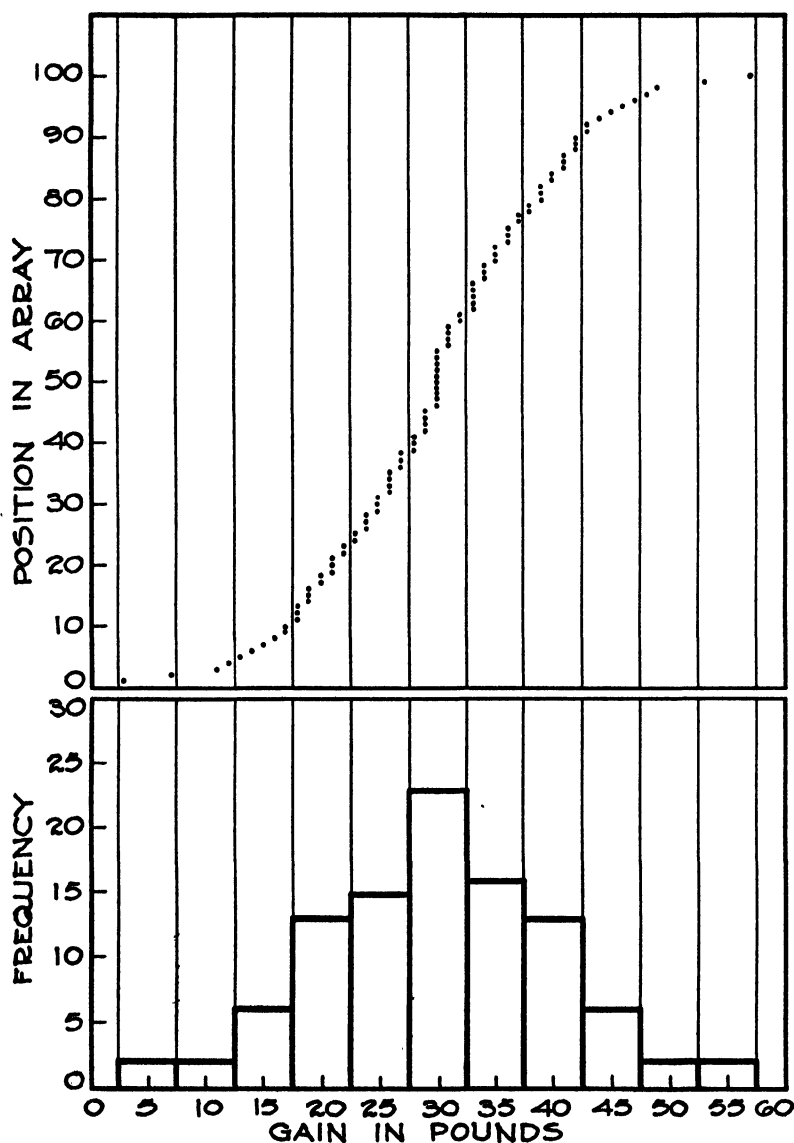


FIG. 3.1—Upper part: Graphical representation of array of 100 normally distributed gains in weight. Lower part: Histogram of same gains. The altitude of a rectangle in the histogram is proportional to the number of dots in the array which lie between the vertical sides.

figure 3.1. Only the class marks are entered in the first row. The class intervals are from 2.5 to 7.5, etc. This is a notably concise method of presenting the data in large samples.

TABLE 3.2
FREQUENCY DISTRIBUTION OF GAINS IN WEIGHT OF 100 SWINE
A normally distributed sample

Class mark (pounds)	5	10	15	20	25	30	35	40	45	50	55
Frequency	2	2	6	13	15	23	16	13	6	2	2

3.3—Random samples from a normal distribution. An easy way to draw random samples from the table of pig gains is to take numbers consecutively from the table of random numbers, table 1.2, then match them with the gains by means of the integers, 00 to 99, in table 3.1. To avoid duplicating the samples of others, start at some randomly selected point in the table of random numbers instead of at the beginning, then proceed upwards, downwards, or crosswise. Suppose you have hit upon the digit, 8, in row 71, column 29. This with the following digit, 3, specifies pig number 83 in table 3.1, a pig whose gain is 40 pounds. Hence, 40 pounds is the first number of the sample. Moving upwards among the random numbers you read the integers 09, 75, 90, etc., and record the corresponding gains from table 3.1, 17, 37, and 43 pounds. Continuing, you get as many gains and as many samples as you wish.

Samples of 10 are suggested. Unlike the procedure in chapter 1, for our present purposes the samples must all be of the same size because the distributions of their statistics change with n . It is well to record the items in columns, leaving a half dozen lines below each for subsequent computations. For your guidance, four samples are listed in table 3.3. The computations below them will be explained as we go along. Draw as many of the samples as you think you can process with the time at your command. It is fortunate if several are working together so that the results of each can be made available to all. Keep the records carefully because you will need them again and again.

It will be noted that each pig gain may be drawn as often as its number appears in the table of random digits—it is not withdrawn from circulation after being taken once. Thus, the sampling is always from the same population, and the probability of drawing any particular item is practically constant throughout the process.

EXAMPLE 3.1—Determine the range in each of your samples of 10. The mean of these ranges is an estimate of $3\sigma = 30$ pounds, table 5.5. Does your average range approximate 30 pounds?

3.4—The distribution of sample means. First add the items in each sample, then put down the mean, $\bar{x}$ (division is by 10). While each of

TABLE 3.3

FOUR SAMPLES OF 10 ITEMS DRAWN AT RANDOM FROM THE PIG GAINS OF TABLE 3.1, EACH FOLLOWED BY STATISTICS TO BE EXPLAINED IN SECTIONS 3.4-3.8

Item Number and Formulas	Sample Number			
	1	2	3	4
1	33	32	39	17
2	53	31	34	22
3	34	11	33	20
4	29	30	33	19
5	39	19	33	3
6	57	24	39	21
7	12	53	36	3
8	24	44	32	25
9	39	19	32	40
10	36	30	30	21
$\bar{x}$	35.6	29.3	34.1	19.1
s^2	169.8	151.6	9.0	112.3
s	13.0	12.31	3.00	10.6
$s_{\bar{x}}$	4.11	3.89	0.95	3.35
t	1.36	-0.18	4.32	-3.25
$t_{.05} s_{\bar{x}}$	9.3	8.8	2.2	7.6
$\bar{x} \pm t_{.05} s_{\bar{x}}$	26.3-44.9	20.5-38.1	31.9-36.3	11.5-26.7

these means is an unbiased estimate of the parameter, 30 pounds, yet there is great variation among them. Make an array of the means of all your samples. If there are enough of them, group them into a frequency distribution like table 3.4.

As indicated, our laboratory means ranged from 19 to 39 pounds, perhaps to the novice a disconcerting variability. Try to imagine doing an experiment resulting in one of these more divergent mean gains instead of the correct value, 30 pounds. Having no information about the population except that furnished by the sample, you would be considerably misled and there is no way to overcome this hazard. One of the objects of the experimental samplings you have done is to acquaint you with the risks involved in all conclusions based on small portions of the aggregate. The investigator seldom knows the parameters of the sampled population: he knows only the sample estimates. He learns to view his experimental data in the light of his experience of sampling error. His judgments involve not only the facts of his sample but all the related information which he and others have accumulated.

The more optimistic are prone to draw satisfaction from the large number of means near the center of the distribution. If this were not characteristic, sampling would not be so useful and popular. The improbability of getting poor estimates produces a sense of security in making inferences. This is proper. With the right balance between confidence and prudence the experienced investigator is seldom a victim of the accidents of sampling.

TABLE 3.4

FREQUENCY DISTRIBUTION OF 511 MEANS OF SAMPLES OF 10 ITEMS RANDOMLY DRAWN FROM THE PIG GAINS OF TABLE 3.1. THEORETICAL NORMAL FREQUENCY FOR COMPARISON

Class Mark (pounds)	Frequency	Theoretical Frequency (For computation, see section 8.8)
19	1	0.20
20	1	0.41
21	0	1.18
22	7	2.71
23	5	5.62
24	10	10.78
25	19	18.60
26	30	29.02
27	41	41.14
28	48	52.07
29	66	61.63
30	72	64.23
31	56	61.63
32	46	52.07
33	45	41.14
34	22	29.02
35	24	18.60
36	12	10.78
37	5	5.62
38	0	2.71
39	1	1.84*
Total	511	511.00

* Including subsequent classes.

The distribution of means in table 3.4 brings to attention two more basic features of the sampling. First, the distribution is itself approximately normal, as you may see by comparing the actual with the theoretical normal frequencies. The theory is that means of samples drawn from a normal population are likewise normally distributed. We may go further and say that means tend to be distributed normally even if the parent population is considerably anormal, a great convenience in practical applications where the exact form of the sampled population is usually unknown. The second feature to be noticed is that the range of the means, great as it is, is only about one-third that of the population of gains in table 3.1. Means are less variable than items, as we have had occasion to remark before. Discussion of this second feature will be resumed in section 3.6 below.

The final characteristic to be noted in the distribution is its mean. This may be calculated by adding all the sample means and dividing by 511, but a shorter method will be explained in section 8.2. The value, 29.87 pounds, is an unbiased estimate of $m = 30$ pounds. Clearly, if one could command large samples such as this he wouldn't bother much about sampling variation. Also, this close approximation throws light

on the meaning of the term, unbiased. As the extent of the sampling has increased, the deviation of the sample mean from the true value has become very small. You remember that an unbiased sample is one whose mean tends toward the population value as n is increased.

3.5—Distribution of variance and standard deviation. s^2 , the average square of the deviations in each of your samples, may be computed as in table 2.1. Shorter methods are explained in sections 5.1 and 5.2, the latter being especially easy if a calculating machine is available. In table 3.3 you will observe that three of these variances overestimate the population value, 100, while the fourth is notably small. Examine any of your samples with unusual variances to learn what peculiarities of the arrays are indicated. This freakish sample 3 in the table has a range of only $39 - 30 = 9$ pounds with not a single item less than the population mean. Did you happen to get a sample with range equal to that of the population?

The distribution of the variances of our 511 samples is displayed in table 3.5. Notice its skewness, with the clustering below the mean and a

TABLE 3.5
FREQUENCY DISTRIBUTION OF 511 VARIANCES OF SAMPLES WITH $n = 10$ DRAWN FROM
THE NORMAL DISTRIBUTION OF TABLE 3.1

Class mark	0	20	40	60	80	100	120	140	160	180	200	220	240	260	280	300	320	340
Frequency	1	11	47	92	93	72	73	42	29	26	11	8	2	1	0	1	1	1

long tail above—somewhat like the chi-square distribution of chapter 1, though less extreme. Despite this, the mean of these variances is 101.5, closely approximating the population value of 100 and verifying the fact that the sample variance is an unbiased estimate.

You may think that we should go ahead with fiducial limits and tests of significance based on the variance distribution. Variance is a measure of dispersion similar to chi-square, and might be used in much the same fashion were it not that the more convenient t has been discovered. This will be discussed in section 3.7.

Our distribution of s , the square root of the variance, in table 3.6 conforms closely to theory. It shows a slight skewness (not so large as

TABLE 3.6
FREQUENCY DISTRIBUTION OF 511 STANDARD DEVIATIONS CORRESPONDING TO THE
VARIANCES OF TABLE 3.5

Class mark	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Frequency	1	2	9	18	58	77	80	71	79	44	41	17	8	3	1	2

that of s^2) as well as a small bias with mean 9.8 pounds, slightly less than $\sigma = 10$ pounds. Thus, even in samples so small as 10 the bias of the standard deviation is negligible in single samples.

EXAMPLE 3.2—Theoretically, 50% of the means in table 3.4 should be smaller than 30 pounds. Assuming that half of the means in the interval with class mark 30 pounds have values less than 30, the sample number of means less than 30 pounds is 264, the remaining 247 being greater. Calculate adjusted chi-square on the hypothesis that in the population the distribution is 1:1. Ans. 0.50.

EXAMPLE 3.3—If half the standard deviations of table 3.6 were expected to be less than 10 pounds, calculate adjusted $\chi^2 = 4.70$ for the sample. This is evidence against a symmetrical distribution in the population.

3.6—Standard deviation of the mean, or standard error. This quantity, one of the most useful in our kit of tools, is calculated from either

$$s_z = s/\sqrt{n} \quad \text{or} \quad s_z = \sqrt{s^2/n},$$

whichever is more convenient. In each of your samples of 10, s_z is an estimate of

$$\sigma_z = \sigma/\sqrt{n} = 10/\sqrt{10} = 3.162 \text{ pounds}$$

In your drawings you will experience the usual sampling variation, but the average of your standard errors should approximate the theoretical 3.162 pounds. Since s_z is slightly biased in the same manner as s , you would get a still more biased average if you calculated the mean of your standard errors. To get the most accurate estimate, add your unbiased variances of the means,

$$s_z^2 = s^2/n,$$

then extract the square root of their mean. In our 511 samples, for example, this average is easily got by taking the average of the variances, 101.5, dividing by 10 and extracting the root, $\sqrt{10.15} = 3.186$ pounds. This close approximation to the population value constitutes sampling evidence that s_z is an estimate of σ_z . But what is the nature of this σ_z of which s_z is an estimate? As indicated in section 2.12, σ_z is the standard deviation of means of samples of n drawn at random from a normal population with standard deviation, σ . Verification of this theorem is found in the distribution of 511 means in table 3.4. Their standard deviation turns out to be 3.178 pounds (in section 8.3 is explained a way of calculating this easier than that of table 2.1), a satisfactory approach to the theoretical 3.162 pounds. If we now combine these two pieces of information, we are prepared for the statement that each random sample affords an estimate, s_z , of the standard deviation of means, or standard error, σ_z . Ordinarily only the single sample is drawn. You have seen from examination of your small samples that the estimate may be disconcertingly divergent, but we have now learned that it is only slightly biased in even small samples of 10. We are next to learn how it leads to fiducial statements and tests of hypotheses.

EXAMPLE 3.4—Four hundred plants of a certain hybrid rye (3) produced their first flowers an average of 70.5 days after planting, the standard deviation being 6.9 days. Estimate the standard deviation of means of similar samples. Ans. $s_{\bar{x}} = 0.34$ day.

EXAMPLE 3.5—Calculate $s_{\bar{x}} = 1.16$ ounces in the following sample of birth weights of 12 pigs: 3, 5, 7, 8, 9, 10, 10, 11, 12, 13, 15, 17 ounces.

EXAMPLE 3.6—During the period from 28 to 70 days of age the mean gain in weight of 16 rats was reported as 138 grams with the standard error, 4 grams. Assuming normal distribution, estimate the range of gains in the sample. (Compare example 2.16). Ans. From 110 to 166 grams.

EXAMPLE 3.7—If you were to draw a sample of 40 items from the pig gains of table 3.1, the standard error of the mean is an estimate of what population parameter? Ans. 1.58 pounds. You can try this conveniently by combining 4 of your samples of 10, taking them just as they came in your original drawing.

EXAMPLE 3.8—Provided you have 10 or more samples of 10, you can easily make a sample of 100. If you used the computational method of section 5.2 your 10 values of SX and SX^2 are quickly combined for calculation of Sx^2 in this sample of 100. $s_{\bar{x}}$, which may then be computed for this sample, is an estimate of what parameter? Ans. 1 pound.

EXAMPLE 3.9—From table 5.5 you may read the average range of samples of 511 as about 6.08σ . Since the means of table 3.4 are drawn from a population in which $\sigma = 10/\sqrt{10} = 3.16$ pounds, calculate the average range expected in samples of 511 and compare the result with the actual range, $39 - 19 = 20$ pounds.

EXAMPLE 3.10—It was reported that a lot of pigs gained at the rate of 1.4 pounds per day with standard error, 0.02 pound per day. The author failed to state the number of pigs in the lot. If you wish to get an idea of this number, you can use the relation, $s_{\bar{x}} = s/\sqrt{n}$ provided you know from experience that the coefficient of variability of fattening swine is about 10%. This would lead to an estimate of the standard deviation as 10% of $1.4 = 0.14$ pound per day. Then, $\sqrt{n} = s/s_{\bar{x}} = 0.14/0.02 = 7$; and finally, $n = 49$ pigs. This value of the coefficient of variation, 10%, is fairly small for fattening swine, but may be reduced still further by efficient experimental methods.

3.7—Distribution of t . You became acquainted with this statistic in sections 2.12 and 2.13, but explanation had to be delayed until now. The formula is

$$t = (\bar{x} - m)/s_{\bar{x}}$$

In your sampling, m is known to be 30 pounds. Contrast this with section 2.13 where the hypothesis was set up that $m = 0$, and where consequently t was equal to $\bar{x}/s_{\bar{x}}$, a value completely determined by the sample. A second contrast is furnished by section 2.12 where the value of m was unknown. The two fiducial limits, l_1 and l_2 , were calculated by solving the equations,

$$t_{.05} = (\bar{x} - l_1)/s_{\bar{x}} \quad \text{and} \quad -t_{.05} = (\bar{x} - l_2)/s_{\bar{x}}$$

in which $t_{.05}$ was a positive number taken from the table, and m was replaced first by l_1 , smaller than $\bar{x}$, and then by l_2 , larger. Clearly, we must know more about this useful quantity, t .

Since $\bar{x}$ and $s_{\bar{x}}$ have already been calculated, the sample value of t may now be got by putting $m = 30$, the resulting formula being

$$t = (\bar{x} - 30)/s_{\bar{x}}$$

Here, t will be positive or negative according as $\bar{x}$ is greater or less than 30 pounds. In the present sampling the two signs are equally likely, so you may expect about half of each. On account of this symmetry the mean of all your t 's should be near zero.

The samples in table 3.3 were selected to illustrate the manner in which large, small and intermediate values of t arise in sampling. A small deviation, $(\bar{x} - m)$, and/or a large standard error tend to make t small. Some striking combinations are put in the table, and you can doubtless find others among your samples.

The distribution of the laboratory sample of t 's is displayed in table 3.7.

TABLE 3.7
SAMPLE AND THEORETICAL DISTRIBUTIONS OF t . SAMPLES OF 10.
DEGREES OF FREEDOM, 9

Interval of t		Sample		Theoretical		
From	To	Frequency	Percentage Frequency	Percentage Frequency	Cumulative	
					One Tail	Both Tails
.....	-3.250	3	0.6	0.5	100.0	
-3.250	-2.821	4	0.8	0.5	99.5	
-2.821	-2.262	5	1.0	1.5	99.0	
-2.262	-1.833	16	3.1	2.5	97.5	
-1.833	-1.383	31	6.1	5.0	95.0	
-1.383	-1.100	25	4.9	5.0	90.0	
-1.100	-0.703	52	10.2	10.0	85.0	
-0.703	0.0	132	25.8	25.0	75.0	
0.0	0.703	126	24.6	25.0	50.0	100.0
0.703	1.100	41	8.0	10.0	25.0	50.0
1.100	1.383	32	6.3	5.0	15.0	30.0
1.383	1.833	18	3.5	5.0	10.0	20.0
1.833	2.262	13	2.5	2.5	5.0	10.0
2.262	2.821	8	1.6	1.5	2.5	5.0
2.821	3.250	2	0.4	0.5	1.0	2.0
3.250		3	0.6	0.5	0.5	1.0
		511	100.0	100.0		

The class intervals in the present table are, as in table 1.4, unequal, adjusted so as to bring into prominence certain useful probabilities in the tails of the distribution. The theoretical percentage frequencies are recorded for comparison with those of the sample. In the last two columns are the cumulative percentage frequencies which make the table convenient for fiducial statements and tests of hypotheses. Examination of the table reveals that 2.5% of all t -values in samples of 10 theoretically fall beyond 2.262, while another 2.5% are smaller than -2.262. Combining these two tails of the distribution, as shown in the last column, 5% of all t 's in samples of 10 lie further from the center than 2.262, which is therefore called the 5% level of t . The 5% point, you will note, beyond

which 5% of *each* tail lies, is ≈ 1.833 . The utility of these and similar markers is to be explained. At the moment, make a distribution of your own sample t 's to be compared with the theoretical distributions in the table.

Cumulative percentage distributions of t for other sample sizes have been worked out and are presented in table 3.8. Because of symmetry, only the positive half of the t 's are listed in the body of the table; that is, only those corresponding to the lower half of table 3.7. Along the top border may be read the cumulative percentages for the two tails combined, while on the left are degrees of freedom, $n - 1$. The boundaries of the class intervals in table 3.7 (0.703, 1.100, 1.383, etc.) may be found in the row specified by 9 degrees of freedom.

3.8—The fiducial statement. The theory of the fiducial interval may now be verified from your sampling. Each sample is the basis of a fiducial statement about the population mean. Put $t = +2.262$, the 5% level for 9 degrees of freedom, in both of the equations given in the foregoing section—

$$2.262 = (\bar{x} - l_1)/s_{\bar{x}} \quad \text{and} \quad -2.262 = (\bar{x} - l_2)/s_{\bar{x}}$$

Solving each,

$$l_1 = \bar{x} - 2.262s_{\bar{x}} \quad \text{and} \quad l_2 = \bar{x} + 2.262s_{\bar{x}}$$

Substituting $\bar{x}$ and $s_{\bar{x}}$ from each sample, the two limits are calculated as illustrated in table 3.3. Finally, for any particular sample, if you say that the population mean lies between the limits l_1 and l_2 you will be right or wrong; which it is may be determined readily because you know the parameter, 30 pounds. The theory will be verified if about 95% of your statements are right and about 5% wrong.

Referring again to table 3.3, sample 1 warrants the statement that the population mean lies between $l_1 = 26.3$ and $l_2 = 44.9$ pounds, and we know in the present instance that this fiducial interval does contain m . On the contrary, samples 3 and 4 were selected to illustrate cases leading to false statements, one because of an unusually divergent mean, the other because of a small standard deviation. Of the 511 laboratory samples, 486 resulted in correct statements about m ; that is, 95.1% of the statements were true. The percentage of false statements, 4.9%, closely approximated the theoretical 5%. Always bear in mind the condition involved in every fiducial statement at the 5% level—it is right unless a one-in-twenty chance has occurred in the sampling.

Practical applications of this theory are by people doing experiments and other samplings without knowledge of the population parameters. When they make fiducial statements they do not know if they are right or wrong—they know only the probability selected. It is possible though improbable that the first five samples you draw in your research career will lead to false statements, not because of unskillful manipulation but

TABLE 3.8
VALUES OF t^*

<i>D.f.</i>	Probability of a Larger Value of t , Sign Ignored								<i>D.f.</i>
	0.5	0.4	0.3	0.2	0.1	0.05	0.02	0.01	
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657	1
2	.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	2
3	.765	.978	1.250	1.638	2.353	3.182	4.541	5.841	3
4	.741	.941	1.190	1.533	2.132	2.776	3.747	4.604	4
5	.727	.920	1.156	1.476	2.015	2.571	3.365	4.032	5
6	.718	.906	1.134	1.440	1.943	2.447	3.143	3.707	6
7	.711	.896	1.119	1.415	1.895	2.365	2.998	3.499	7
8	.706	.889	1.108	1.397	1.860	2.306	2.896	3.355	8
9	.703	.883	1.100	1.383	1.833	2.262	2.821	3.250	9
10	.700	.879	1.093	1.372	1.812	2.228	2.764	3.169	10
11	.697	.876	1.088	1.363	1.796	2.201	2.718	3.106	11
12	.695	.873	1.083	1.356	1.782	2.179	2.681	3.055	12
13	.694	.870	1.079	1.350	1.771	2.160	2.650	3.012	13
14	.692	.868	1.076	1.345	1.761	2.145	2.624	2.977	14
15	.691	.866	1.074	1.341	1.753	2.131	2.602	2.947	15
16	.690	.865	1.071	1.337	1.746	2.120	2.583	2.921	16
17	.689	.863	1.069	1.333	1.740	2.110	2.567	2.898	17
18	.688	.862	1.067	1.330	1.734	2.101	2.552	2.878	18
19	.688	.861	1.066	1.328	1.729	2.093	2.539	2.861	19
20	.687	.860	1.064	1.325	1.725	2.086	2.528	2.845	20
21	.686	.859	1.063	1.323	1.721	2.080	2.518	2.831	21
22	.686	.858	1.061	1.321	1.717	2.074	2.508	2.819	22
23	.685	.858	1.060	1.319	1.714	2.069	2.500	2.807	23
24	.685	.857	1.059	1.318	1.711	2.064	2.492	2.797	24
25	.684	.856	1.058	1.316	1.708	2.060	2.485	2.787	25
26	.684	.856	1.058	1.315	1.706	2.056	2.479	2.779	26
27	.684	.855	1.057	1.314	1.703	2.052	2.473	2.771	27
28	.683	.855	1.056	1.313	1.701	2.048	2.467	2.763	28
29	.683	.854	1.055	1.311	1.699	2.045	2.462	2.756	29
30	.683	.854	1.055	1.310	1.697	2.042	2.457	2.750	30
35						2.030		2.724	35
40						2.021		2.704	40
45						2.014		2.690	45
50						2.008		2.678	50
60						2.000		2.660	60
70						1.994		2.648	70
80						1.990		2.638	80
90						1.987		2.632	90
100						1.984		2.626	100
125						1.979		2.616	125
150						1.976		2.609	150
200						1.972		2.601	200
300						1.968		2.592	300
400						1.966		2.588	400
500						1.965		2.586	500
1000						1.962		2.581	1000
∞	.67449	.84162	1.03643	1.28155	1.64485	1.95996	2.32634	2.57582	∞

* Partly from table IV of "Statistical Methods for Research Workers" with permission of the author, R. A. Fisher, and of his publishers, Oliver and Boyd, Edinburgh.

through an unlucky quirk of chance. Let us hope that the wise counsel of a more experienced leader would guide you around such a sampling pitfall.

I trust you now have a clear view of the nature of the fiducial inference, and that you will make use of it with a happy combination of caution and confidence. In the following sections we shall study the third inference involving certain hypotheses about populations.

EXAMPLE 3.11—Using the sample frequencies of table 3.7, test the hypothesis (known to be true) that the t -distribution is symmetrical in the sense that half the population frequency is greater than zero. Ans. $\chi^2 = 1.13$.

EXAMPLE 3.12—From table 3.7 it is seen that $3 + 4 + 5 + 8 + 2 + 3 = 25$ samples have $|t| > 2.262$. Test the hypothesis that 5% of the population values are greater than 2.262. Ans. $\chi^2 = 0.0001$.

EXAMPLE 3.13—In table 3.7, accumulate the sample frequencies in both tails and compare their percentage values with those in the last column of the table.

EXAMPLE 3.14—During the fall of 1943, approximately one in each 1,000 city families of Iowa (cities are defined as having 2,500 inhabitants or more) was visited to learn the number of quarts of food canned. The average for 300 families was 165 quarts with standard deviation, 153 quarts. Calculate the 95% fiducial limits. Ans. 165 ± 17 quarts.

EXAMPLE 3.15—The 1940 census reported 312,000 dwelling units (roughly the same as families) in Iowa cities. From the statistics of the foregoing example, estimate the number of quarts of food canned in Iowa cities in 1943. Ans. 51,500,000 quarts with 95% fiducial limits, 46,200,000 and 56,800,000 quarts.

EXAMPLE 3.16—Compute the 99% fiducial limits of the mean in example 3.14. Ans. 165 ± 23 quarts.

3.9—Differences. Test of null hypotheses. Beginning with section 1.6 and again with section 2.13 you learned how to test useful hypotheses pertaining to sampled populations. It is time to examine the latter section more in detail, using the familiar technique of drawing samples from the hypothetical population. In practice, the hypothesis most often set up is that there is no difference between the two populations sampled; that is, that the contrasted members of the pairs are drawn from the same source. A convenient model is a normally distributed population of differences with mean equal to zero. We can easily sample such a population by taking advantage of this theorem: if items are drawn at random from a normal population, then randomly paired, the differences between pairs are normally distributed with zero as mean. Merely copy down your samples in pairs, as in table 3.9, subtract each item in the second from the corresponding item in the first, and use the differences as the new samples. Naturally, the samples must be taken just as they came from the original drawing, without selection as to which shall be first and which second, and there must be no rearrangement of the items. Two of the pairs in the table were selected from our 511 samples to illustrate features to be commented on below. The third is a made-up sample illustrating a theorem that will be needed later.

These two facts will be noticed right away in the sampling: (i) the

TABLE 3.9
TWO SAMPLES OF DIFFERENCES DRAWN AT RANDOM FROM TABLE 3.1,
AND A MADE-UP SAMPLE

$$m = m_1 - m_2 = 0. \quad \sigma_1^2 = \sigma_2^2 = 100. \quad \sigma^2 = 200$$

1. Random			2. Random			3. Selected		
Paired Items	Differences		Paired Items	Differences		Paired Items	Differences	
X_1	X_2	X	X_1	X_2	X	X_1	X_2	X
39	17	22	19	33	-14	32	22	10
34	29	5	14	30	-16	31	17	14
22	30	-8	57	41	16	28	34	-6
27	36	-9	34	42	-8	24	24	0
33	41	-8	39	33	6	44	12	32
42	30	12	34	21	13	53	29	24
36	3	33	39	36	3	9	23	-14
24	23	1	13	33	-20	35	37	-2
25	38	-13	39	33	6	33	21	12
29	30	-1	23	22	1	31	11	20
$\bar{x}$	3.4			-1.3			9	
s^2	222.9			155.6			207.4	
s_x^2	22.29			15.56			20.74	
s_z^2	4.72			3.94			4.55	
t	0.72			-0.33			1.98	

differences, both in size and number, are almost equally divided between pluses and minuses, resulting in means falling with equal likelihood above and below zero; and (ii) the variation of the differences is much greater than that of the original population. The reason for this second fact is apparent when you observe that the possible range of differences is from $3 - 57 = -54$ pounds to $57 - 3 = 54$ pounds, a range of $54 - (-54) = 108$ pounds, twice the range of the gains in table 3.1. You will not be surprised, then, by these theorems: (i) the variance of random differences, σ^2 , is double that of the original population; and (ii) the variance of each sample of differences, s^2 , is an unbiased estimate of $2\sigma_1^2$, ($= 2\sigma_2^2$). So if you average the variances of all your samples of differences, the result should be not far from 200, twice the variance of the pig gains. I had 144 such variances with the mean, 207.9. You are doubtless impressed, perhaps startled, by the variability of s^2 in your samples.

You are now prepared for another theorem: not only the differences but also the *means of the differences* are normally distributed about $m = 0$, the variance of the means being $\sigma_x^2 = 2\sigma_1^2/n$. Again, the variance of each sample mean, s_x^2 , is an unbiased estimate of σ_x^2 . In verification, I found the average of my 144 mean differences to be 0.51 pound with variance, 22.2, instead of the expected $m = 0$ and $\sigma_x^2 = 20$.

This variance, double that in the original population, may be disconcerting to the many who expect to set up experiments for evaluating

differences. Being confronted by such great variation, how can they hope to get means with fiducial intervals small enough to be of practical use? Obviously, larger samples would do it, but such may not be feasible. Some devices must be found for reducing the error encountered, or else random pairing must be avoided whenever possible. Both alternatives are to be considered in the next section. At the moment we must look at the values of t in our samples of differences. Since m is now zero, the formula is

$$t = \bar{x}/s_x,$$

as in section 2.13. The calculation is illustrated in table 3.9. Here is an interesting new feature. Since the differences have greater variability than the original gains, both $\bar{x}$ and s_x have wider ranges; but the ratio, t , is distributed exactly as in table 3.7, a fact substantiated by the distribution of table 3.10. Clearly, these 144 t 's might be added to the previous

TABLE 3.10
DISTRIBUTION OF 144 VALUES OF t IN SAMPLES OF DIFFERENCES
RANDOMLY DRAWN FROM TABLE 3.1; $n = 10$

Class Interval*	Frequency	Percentage Frequency	Theoretical Percentage Frequency
..... -2.821	0	0.00	1
-2.821 -1.833	6	4.17	4
-1.833 -1.100	11	7.64	10
-1.100 0.000	48	33.33	35
0.000 1.100	55	38.20	35
1.100 1.833	13	9.03	10
1.833 2.821	10	6.94	4
2.821	1	0.69	1
	144	100.00	100

* Because this sample is small, each pair of intervals in table 3.7 is here combined into one.

511, making a sample of 655 values all randomly and independently drawn from a single population of t . This justifies the use of the same table of t for samples of gains and for samples of differences. For any one size of sample the distribution of t is contingent only upon normality in the population sampled, irrespective of m or σ . This is one reason for the utility of the statistic.

In your samples of differences, did you get any values of t larger (in absolute value) than the 5% level, 2.262? Among my 144 samples I got 10 instead of the expected 7. One was greater than 3.250, the 1% level. You are well aware by now that if these were experimental results, the null hypothesis might have been rejected in these 10 samples though it was, in fact, true.

EXAMPLE 3.17—Collins, Flint, and McLane (2) investigated the effect of a small electric current on the growth of corn seedlings. The seedlings were grown in pairs of boxes, one of each pair treated, the other not. The differences in elongation, treated-untreated, were as follows: 6.0, 1.3, 10.2, 23.9, 3.1, 6.8, -1.5, -14.7, -3.3, 11.1 mm. Calculate $t = 1.33$.

EXAMPLE 3.18—An agronomist interested in the effect of superphosphate on corn yield tried adding the fertilizer to a treatment of manure and lime. Five pairs of plots were tested. The plots with superphosphate yielded 20, 6, 4, 3, and 2 bushels per acre more than their parallels. Was the value of superphosphate demonstrated?

EXAMPLE 3.19—The foregoing experiment was tried again. The second time, the plots receiving superphosphate yielded 5, 6, 4, 3, and 2 bushels per acre more than their parallels. Was this better evidence of the value of superphosphate than the former? If you think merely of the samples, the answer seems silly. If you fix your attention on the populations, the results are reasonable. The sample estimate indicates that the first population is highly variable, so that small samples drawn from it are unreliable.

EXAMPLE 3.20—The following are the yields in bushels per acre of 22 corn crops from Iowa farms on Carrington loam. The data were collected by Boatman and Baker in 1918–1924. The farms are in the locality of the towns indicated. Is the mean difference significant? Do you suppose the lime directly affected the yield of corn?

Locality	Eldora		Gruver		Jesup		Scranton
Year	1919	1920	1923	1924	1921	1922	1924
Fertilized with manure only	50	62	38	60	73	66	43
Fertilized with manure and lime	52	66	40	64	78	71	46

Truesdale, 2 farms							Independence		Maynard	Waverly, 3 farms						
1918	1919	1920	1921	1922	1924		1921	1924	1923	1918	1920	1922	1922	1923	1923	
44	57	57	40	62	24		64	58	54	54	89	58	58	40	25	
46	58	59	42	64	32		69	59	60	57	88	60	62	57	30	

EXAMPLE 3.21—It is reported that, in a series of 15 variety trials, the difference in plot yields, $A - B$, ranged from -6 to 16 bushels per acre. Can you estimate $t = 3.2$ for testing the significance of the mean difference between A and B ?

EXAMPLE 3.22—In reporting the results of his research, an investigator stated the mean difference and standard error of his differences but omitted the sample size. He did say, however, that his t -value was greater than the 1% level, 2.861. Would you deduce the size of his sample as 20 differences?

EXAMPLE 3.23—Two lots of 10 rats each, paired as usual, were fed rations producing a mean difference of 4.4 grams in weight, with standard deviation 9.4 grams. Since the difference is not significant, the question is raised as to whether more rats might demonstrate a significant superiority of one of the rations. Of course, the population gains corresponding to the two rations may not be different. Assuming, however, that they may be, and assuming further that a trial with more rats will produce a mean difference of not less than 4.4 grams, and a standard deviation of not

more than 9.4 grams, find the number of pairs necessary to make $t = 2$ (section 2.14).
Ans. About 19 pairs.

EXAMPLE 3.24—To get a clear idea of the 5% values of t in table 3.8, make a graph with degrees of freedom along the horizontal axis and t plotted against them vertically. The curve through such points will fall away rapidly at first, then more slowly, approaching the value, $t = 1.96$, as a limit. If carefully drawn, this graph may be used for interpolating t -values between those given in the table.

EXAMPLE 3.25—For an interesting account of some of the early testing of experimental data from paired individuals, read Fisher's *Design of Experiments* (4), chapter 3. It is there shown that even Darwin and Galton were uncertain of the principles governing this test of significance.

All the facts necessary for reading the earlier sections of chapter 4 have been presented. For a brief first course, therefore, the remainder of this chapter may be omitted.

3.10—Reducing the error involved in sampling. Having observed the large variation in random differences, we felt that something must be done to avoid such inaccuracy in experimentation. Extending the sampling to great numbers of pairs would decrease s_x to any specified value, but this process would be expensive. Furthermore, since practical samplings are often from populations restricted in size, any great expansion of n may be impossible. Therefore, samplers have sought other methods for increasing accuracy.

Three additional ways of reducing sampling errors are in common use: (i) selecting or developing populations in which variation is known to be small, (ii) subdividing the aggregates into sub-populations or *strata* of similar individuals, and (iii) utilizing information contained in related variates. These methods may be used in all combinations. Some details follow immediately, but most of them are distributed throughout the remainder of the book.

It is sometimes practicable to get information by sampling populations whose variation is small. For example, one might decide to restrict his inquiry about radio program listeners to women in a certain city, questioning only those who are housekeepers with one or more children and whose family income is between, say, \$2,000 and \$3,000. The variable being measured, such as amount spent for toilet soap, might have little range within this restricted population. Clearly, no specific information would be got about other parts of the aggregate, but the information about the population sampled should be relatively accurate. Again, for certain problems in nutrition of rats it seems adequate to sample a highly inbred colony whose range in 28-day weight, for example, is half that resulting from random mating. Standard errors for means of n such animals would be half those in the randomly mated population; consequently, fiducial intervals would be decreased by the same fraction while sample values of t would be doubled.

The facts just stated can be verified by sampling from the normal population represented by table 3.11. This population has $\sigma = 5$ pounds instead of the 10 pounds of table 3.1, the range also being half the previous value. Since the facts seem obvious, it will scarcely pay you to draw a

TABLE 3.11

ARRAY OF 100 GAINS HAVING HALF THE RANGE AND STANDARD DEVIATION AS THAT IN TABLE 3.1. DISTRIBUTION IS APPROXIMATELY NORMAL WITH $m = 30$ POUNDS AND $\sigma = 5$ POUNDS

Item No.	Gain	Item No.	Gain	Item No.	Gain	Item No.	Gain
00	17	25	27	50	30	75	33
01	19	26	27	51	30	76	34
02	20	27	27	52	30	77	34
03	21	28	27	53	30	78	34
04	22	29	27	54	31	79	34
05	22	30	27	55	31	80	34
06	22	31	28	56	31	81	34
07	23	32	28	57	31	82	35
08	23	33	28	58	31	83	35
09	23	34	28	59	31	84	35
10	23	35	28	60	31	85	35
11	24	36	28	61	31	86	35
12	24	37	29	62	32	87	36
13	24	38	29	63	32	88	36
14	25	39	29	64	32	89	36
15	25	40	29	65	32	90	36
16	25	41	29	66	32	91	37
17	25	42	29	67	32	92	37
18	26	43	29	68	33	93	37
19	26	44	29	69	33	94	38
20	26	45	30	70	33	95	38
21	26	46	30	71	33	96	39
22	26	47	30	72	33	97	40
23	26	48	30	73	33	98	41
24	27	49	30	74	33	99	44

lot of samples from this table—its real use will come later. It is probably sufficient to look at the distribution of 400 means of 10 items drawn in the laboratory, table 3.12. The range is seen to be about half that in table 3.4. The average of the means in table 3.12 is 30.0 pounds, while their standard deviation is 1.57 pounds instead of the expected

$$\sigma_{\bar{x}} = \sigma/\sqrt{n} = 5/\sqrt{10} = 1.58 \text{ pounds}$$

Obviously, this scheme of reducing sampling variation by sampling from small-ranged populations is effective. Unfortunately, there are not many circumstances in which the method is available. Usually the information desired is about existent populations over whose variation the sampler

TABLE 3.12

FREQUENCY DISTRIBUTION OF MEANS OF SAMPLES OF 10 ITEMS DRAWN AT RANDOM FROM THE NORMAL DISTRIBUTION IN TABLE 3.11

Class mark (pounds)	26	27	28	29	30	31	32	33	34
Frequency	5	20	43	80	99	85	48	18	2

has no control, and the less variable population may not be representative of the other.

The second method of reducing error variation, by sub-division of the aggregate into relatively homogeneous strata, has been found useful in a surprising variety of situations. Most sampling and experimental designs employ it either exclusively or in connection with other devices. The reader will be reminded of it many times as he pursues this text, and later will be referred to more specialized discussions. An instance of this second method is the sampling from pairs of half leaves described in section 2.13, where the concentration of a virus was being examined. The leaf areas to be inoculated with the two preparations were chosen adjacent because such had been found to react more nearly alike than areas on different leaves. Therefore, if one half of each leaf is rubbed with preparation number 1 and the other half with number 2, the lesion differences are expected to vary less than they would if the two preparations were applied to half leaves selected by chance from different plants.

Was this design successful in the tobacco virus experiment? For an answer, compare the experimental variance, 18.57 as calculated in table 2.2, with an estimate of what it would have been in a completely random arrangement. While the correct way of making this estimate is more conveniently presented in section 11.5, a good approximation is got by calculating the lesion variance on the 8 half-leaves inoculated with the first preparation, then on the 8 with the second, and finally averaging the two. Clearly, the variation due to treatment is not included in this average, but only that arising from the differences among half-leaves treated alike. The two variances, one computed from each of the two columns of half-leaf lesions in table 2.2, are 66.86 and 24.57, the average being 45.71. To compare this with the actual experimental variance, one must remember that 18.57, the variance of differences, is an estimate of $2\sigma^2$. The proper comparison, then, is $18.57/2 = 9.28$ with 45.71. Thus, the variance involved in evaluating this experiment is only about 20% of that which would have been encountered with random selection of half-leaves. This means that, for equal accuracy, five times the number of plants would be required if the pairing of contiguous half-leaves were not practiced. Hence, it is said that the efficiency of this experiment was five times as great as it would have been with random choice. Another way of saying the same thing is that the information per plant was five times that in the nonstratified experiment. (In chapter 11 you will find that the unbiased estimate of this factor is 4.66 instead of 5).

As will be emphasized later, the efficacy of this method of stratification depends on foreknowledge of the behavior of experimental material, and this can be gained only by experience. No advantage is to be expected unless the members of a pair tend to perform in a similar fashion when treated alike. If these members differ as much as the pairs differ, then the stratification is no better than random pairing.

Does stratification imply a desertion of the fundamental principle of randomness? Not at all. It is merely a device for utilizing any knowledge one may have of outcome in order to exclude from the estimate of error that variation which he can anticipate. In the virus experiment two sources of variation had been learned from experience: (i) the differences from plant to plant, and (ii) the differences among leaves on the same plant. Such variability was avoided in estimating experimental error by trying the two preparations on halves of the same leaf. Randomness was attained by allotment of the two preparations to the right and left halves by some such mechanical device as the toss of a coin or a table of random digits. The variation entering the estimate of error was only that due to random differences in the reactions of two areas of the same leaf.

The third method of controlling error variation, by use of additional information from the measurement of related variates, will be presented in chapter 6.

EXAMPLE 3.26—In example 2.27 were given the biological values of raw and roasted peanuts as determined in 10 pairs of rats. The standard deviation of the differences was 7.41, so that the experimental variance per rat is estimated as $(7.41)^2/2 = 27.5$. Calculate the average variance among the rats in the two groups. Ans. 26. It appears then that, so far as biological values are concerned, the pairing was less efficient than random assortment.

EXAMPLE 3.27—If the ninth pair, 44, 62, had not turned up in example 2.27, computation will show that the paired sample would have been judged 9 times as efficient as the unpaired. Since such radical changes in attitude depend on one pair of rats, the investigator learns to be cautious.

3.11—Summary. By experience you have verified a number of useful theorems about sampling from normally distributed populations of measurement variates. Means of samples of n are normally distributed with mean, m , but with standard deviation, $\sigma/\sqrt{n}$. Each random sample yields unbiased estimates of m , σ^2 and σ_x^2 .

The sample, together with some hypothesis about m , yields the statistic, $t = (\bar{x} - m)/s_x$, symmetrically distributed about the mean, zero. With a table of the distributions of t , one may make fiducial inferences about the mean of the population from which the sample is drawn.

Random differences from a single population with variance, σ_1^2 , are normally distributed about zero mean with variance, $\sigma^2 = 2\sigma_1^2$, means of such differences being similarly distributed with $\sigma_x^2 = 2\sigma_1^2/n$. The sample value of $t = \bar{x}/s_x$ furnishes a test of the hypothesis that $m = 0$.

There are three methods of avoiding the variability of random differences: (i) selecting less variable populations, (ii) subdividing the population into strata known to be less variable, and (iii) measuring a related variate.

REFERENCES

1. A. A. BRYAN. Iowa Agricultural Experiment Station Research Bulletin 163 (1933).
2. G. N. COLLINS, L. H. FLINT and J. W. McLANE. Journal of Agricultural Research 38:585 (1929).
3. PEDRO A. DAVID. Iowa State College Journal of Science, 5:285 (1931).
4. R. A. FISHER. Design of Experiments. Oliver and Boyd, London (1942).
5. W. B. MERCER and A. D. HALL. Journal of Agricultural Science, 4:107 (1911).
6. GUSTAV A. WIEBE. Journal of Agricultural Research, 50:331 (1935).

SOME TEXTS ON MATHEMATICAL STATISTICS

- JOHN F. KENNEY. Mathematics of Statistics. Two parts. D. Van Nostrand Company, Inc., New York (1939).
- PAUL R. RIDER. An Introduction to Modern Statistical Methods. John Wiley and Sons, Inc., New York (1939).
- S. S. WILKS. Mathematical Statistics. Princeton University Press, Princeton, New Jersey (1943)

Chapter 4

THE COMPARISON OF TWO GROUPS

4.1—Introduction. Proposed in the foregoing chapter were three devices for decreasing or avoiding the disagreeable variation that may be found in random differences. If no one of these schemes is feasible it may be wise, in making comparisons, to refrain from the pairing of individuals, merely assigning them at random to two groups (lots) and applying one treatment to each group *en masse*. It will then be only the group means that are compared, not the individual measurements. The variation among the individual measurements, however, should be the basis for an estimate of experimental error. An example will illustrate this type of comparison.

4.2—An experiment comparing two groups having equal numbers of individuals. Using male chicks, Breneman (2) contrasted the 15-day mean comb weight of two lots, one receiving sex hormone A (testosterone) and the other C (dehydroandrosterone). Day-old chicks, eleven in number, were assigned at random to each of the treatments. The two lots, caged together, were distinguished by red and purple head-stains. The individual comb weights are recorded in table 4.1.

TABLE 4.1
WEIGHT OF COMBS* (15TH DAY) OF MALE CHICKS GIVEN HORMONES A AND C

Hormone	Number of Chicks	Weight of Comb (milligrams)											Mean
A	11	57	120	101	137	119	117	104	73	53	68	118	97
C	11	89	30	82	50	39	22	57	32	96	31	88	56

* Slight alterations of the original data have been made to simplify computations.

Since there was no basis for pairing the birds, no individual differences are computed. It is the difference of two means, $\bar{x} = 97 - 56 = 41$ mg., that is an unbiased estimate of the difference between the means in two populations of comb weights characterized by the hormones administered.

To arrive at an estimate of the variation of $\bar{x}$, one naturally turns to the sampling experiments of the previous chapter. He may use the means of samples drawn from the population of pig gains, table 3.1. If these means are paired at random, as in table 4.2, the differences will vary

TABLE 4.2
RANDOMLY ASSORTED MEANS OF SAMPLES OF 10 DRAWN FROM THE DISTRIBUTION OF
TABLE 3.1, TOGETHER WITH DIFFERENCES
Pounds

Means			Difference		
$\bar{x}_1$	$\bar{x}_2$		$\bar{x}_1$	$\bar{x}_2$	$\bar{x}$
26	25	1	31	30	1
28	26	2	33	29	4
32	37	-5	33	24	9
31	30	1	37	32	5
31	35	-4	26	31	-5
29	36	-7	34	27	7
35	29	6	24	29	-5
32	34	-2	29	33	-4
29	33	-4	26	30	-4
33	28	5	38	28	10
30	33	-3	24	32	-8
33	27	6	30	31	-1
32	29	3	31	27	4
22	25	-3	32	26	6
21	30	-9	35	28	7
29	26	3	30	28	2
28	34	-6	31	33	-2
31	36	-5	26	33	-7
29	30	-1	29	22	7
28	30	-2	27	28	-1
29	32	-3	32	30	2
30	34	-4	29	28	1
26	30	-4	26	27	-1
28	27	1	25	27	-2
30	33	-3	33	29	4

Variance of differences, 22.42. Standard deviation of differences, 4.7 pounds.

by an amount that will indicate the variance desired. Computation in the usual fashion results in $s_x^2 = 22.42$, somewhat more than twice the variance of means which was encountered in the sampling of section 3.6, and almost the same as the variance of mean differences found in section 3.9. Indeed, the theory is the same whether the mean of differences or the difference of means is examined. In both processes we are estimating $2\sigma^2/n = 20$. This is the clue we need to solve the problem of group comparisons. If we estimate the variance, σ^2 , of the sampled population, we can then estimate the variance of the differences between sample means. As always in actual sampling these estimates are furnished by the sample statistics.

Turning then, to table 4.3, two features may be noted: (i) each sample furnishes an estimate of variance, and (ii) the two estimates are combined or *pooled* by adding the two sums of squares, then dividing by the total degrees of freedom. Use of this pooled variance with *d.f.* = $2(n - 1)$ enables us to take advantage of the *t*-distribution, because *t* calculated as in table 4.3 is distributed in the same manner as heretofore (table 3.8).

TABLE 4.3
SUMMARY OF STATISTICS FOR COMPARISON OF TWO GROUPS OF EQUAL SIZE.
CHICK DATA OF TABLE 4.1

Hormone	Number of Chicks	Degrees of Freedom	Mean Comb Weight (milligrams)	Sum of Squares
A	11	10	97	8,472
C	11	10	56	7,748
Sum = 20			Difference = $\bar{x} = 41$	Sum = $Sx^2 = 16,220$

$$\text{Pooled variance} = s^2 = 16,220/20 = 811$$

$$s_{\bar{x}} = \sqrt{2s^2/n} = \sqrt{2(811)/11} = \sqrt{147.5} = 12.14 \text{ mgm.}$$

$$t = 41/12.14 = 3.38. \quad P < 1\%$$

The remaining calculations follow familiar courses. The variance of means is estimated by $s^2/n = 811/11$, while that of differences between means is double this, or $s_{\bar{x}}^2 = 2s^2/n = 147.5$.

We are now ready for a test of significance. This experiment was set up partly to learn if the hormones differentiated the chicks into two separate populations with contrasting means. To get evidence on this, we make the hypothesis that the populations have the same m , then submit this hypothesis to test in the usual manner. We are asking if the sample difference, $\bar{x} = 41$ mg., may be no more than sampling variation from the hypothetical difference, zero. We have $\bar{x} = 41$ mg., $m = 0$, $s_{\bar{x}} = 12.14$ mg., and

$$t = 41/12.14 = 3.38$$

For 20 degrees of freedom, table 3.8 indicates small chance of a larger value of t if sampling is really from the hypothetical population. The evidence is convincing that these hormones differentiate comb weights into two populations with unequal means.

After the reader has mastered the argument leading to table 4.3 he may find it convenient to summarize the calculations by use of this formula:

$$t = \bar{x} \sqrt{\frac{n(n-1)}{Sx^2}}$$

in which $\bar{x}$ designates the difference between the two group means and Sx^2 is the *pooled* sum of squares. Substituting the chick data, $\bar{x} = 97 - 56 = 41$ mg., $n = 11$, and $Sx^2 = 8,472 + 7,748 = 16,220$. Then

$$t = 41 \sqrt{\frac{(11)(10)}{16,220}} = 3.38, \text{ as before.}$$

Having reached the conclusion that there are two population means instead of one, it is informative to set fiducial limits to their difference, of which 41 mg. is an unbiased estimate. Since $t_{.05} = 2.086$ for $d.f. = 20$, the 95% limits are

$$41 - (2.086)(12.1) \quad \text{and} \quad 41 + (2.086)(12.1),$$

that is, 16 and 66 mg. If the experimenter states that the difference between the means of the two populations is within these limits he will be right unless a one-in-twenty chance has come off.

EXAMPLE 4.1—From the normally distributed pig gains of table 3.11, two samples of 10 were drawn at random. With minor alterations to make calculation easier, the gains in weight (pounds) were:

Sample 1: 35, 27, 31, 36, 40, 32, 29, 27, 30, 33

Sample 2: 24, 29, 26, 31, 32, 28, 26, 35, 32, 27

The mean difference is 3 pounds with pooled variance, 14.44. Calculate $t = 1.76$. Note that the variance, 14.44, is an estimate of the population variance, 25.

EXAMPLE 4.2—In the foregoing example, the pooled sum of squares was 260. Use the formula, $t = \bar{x}\sqrt{n(n-1)/Sx^2}$, to calculate t .

EXAMPLE 4.3—Lots of ten bees were fed two concentrations of syrup, 20% and 65%, at a feeder half a mile from the hive (7). Upon arrival at the hive their honey sacs were removed and the concentration of the fluid measured. In every case there was a decrease from the feeder concentration. The decreases were: from the 20% syrup, 0.7, 0.5, 0.4, 0.7, 0.5, 0.4, 0.7, 0.4, 0.2, and 0.5; from the 65% syrup, 1.7, 2.8, 2.2, 1.4, 1.3, 2.1, 0.8, 3.4, 1.9, and 1.4%. Show that $t = 5.6$ if $m = 0$. There is little doubt that, under the experimental conditions imposed, the concentration during flight decreases more with the 65% syrup. See section 4.6 for further discussion.

EXAMPLE 4.4—In a trial of the effect of supplementary foodstuffs on the iodine number of back fat, four pigs were placed in each of two lots. In the lot fed tankage the iodine numbers were 65.5, 65.3, 67.1, and 73.7; while in the lot where 5% of soybeans was substituted for the corresponding amount of corn, they were 71.5, 76.4, 72.0, and 72.1. Show that the difference between means, 5.1, with $s_x = 2.28$, is not quite significant. Ans. $t = 2.24$.

EXAMPLE 4.5—Four determinations of the pH of Shelby loam were made with each of two types of glass electrode (5). With a modified quinhydrone electrode, the readings were 5.78, 5.74, 5.84 and 5.80; while with a modified $Ag/AgCl$ electrode, they were 5.82, 5.87, 5.96, and 5.89. With the hypothesis that $m = 0$, calculate $t = 2.66$.

EXAMPLE 4.6—Fifteen kernels of mature Iodent corn were tested for crushing resistance. Measured in pounds the resistances were: 50, 36, 34, 45, 56, 42, 53, 25, 65, 33, 40, 42, 39, 43, 42. Another batch of 15 kernels was tested after being harvested in the dough stage: 43, 44, 51, 40, 29, 49, 39, 59, 43, 48, 67, 44, 46, 54, 64. Test the significance of the difference between the two means. Ans. $t = 1.38$.

EXAMPLE 4.7—When yearling steers are shipped in from the range for fattening, so little is known of their previous history that estimating outcome is an uncertain process. They may be just as well placed in lots at random. Culbertson and Hammond (4) report the average daily gains of two lots of steer calves, one (I) having a supplement of linseed oil meal, and a second (V) receiving a light allowance (1 pound per steer daily) of whole soybeans. The rates of gain were: (I) 1.95, 2.17, 2.06, 2.11, 2.24, 2.52, 2.04, 1.95; (V) 1.82, 1.85, 1.87, 1.74, 2.04, 1.78, 1.76, 1.86 pounds per day. Calculate the difference of the means, 0.29 pound per day, $s = 0.148$ pound per day, and $t = 3.92$.

EXAMPLE 4.8—In reading reports of researches it is sometimes desirable to supply a test of significance which was not considered necessary by the author. As an example, Smith (10) gave the mean yields and standard errors of two crosses of maize as 8.84 ± 0.39 and 7.00 ± 0.18 grams. Each mean was the average of five replications. We wish to determine if the mean difference is significant. It is necessary to retrace the original computations till each sum of squares is reached. Starting with $s_2 = 0.39$, $s_2^2 = 0.1521$, $s^2 = 0.7605$, and $Sx^2 = 3.042$. Similarly, in the second cross, $Sx^2 = 0.648$. Pooling these two sums of squares with $d.f. = 8$, the value of t may be calculated. Ans. $t = 4.29$.

EXAMPLE 4.9—Derive the formula, $t = \bar{x}\sqrt{n(n-1)}/Sx^2$, starting with the rules of calculation given in connection with table 4.3.

4.3—Difficulties met in experiments with groups. The foregoing test of the null hypothesis, $m = 0$, involves the subsidiary hypothesis that the comb weights in the two populations are normally distributed with common variance.

As usual, the accuracy of the probabilities is little affected by even considerable departures from normality. The matter of population variance will be discussed in section 4.6. More immediate is the question as to whether the estimate of variance, s^2 , correctly evaluates the variation, aside from treatment effects, to which the experimental means are subject. In the present instance this is aimed at by choosing chicks of one breed and sex, all of the same age, then housing them together. Presumably no extraneous environmental factors differentiated the lots, only the hormone treatments being responsible for the population difference in means. The designs of many group comparisons leave these matters in doubt. If lots are kept in separate cages, or if two-groups of pots are affected by positional differences in the greenhouse, the indicated population difference may be a combination of treatment and other environmental effects and s^2 may be a biased estimate of σ^2 . The treatment differences are then said to be *confounded* with the others, and there may be no means of distinguishing them. In these circumstances, statistical evidence that the population averages are different may have little bearing on the problem of possible treatment effects. Although the statistics may lead correctly to rejection of the null hypothesis, and although this hypothesis may, in fact, be untrue, yet the treatments themselves may have no effects, or may tend to cause a difference opposite to that in the sample. Furthermore, there may be no way to estimate separately the amounts or directions of the two kinds of environmental effects, treatment and extraneous, the population difference being compounded of such effects in unknown proportions. If the experiment is conducted so as to avoid such confusion, σ^2 will be estimated by s^2 and real treatment effects may be expected to lead to significant differences.

The reader who omitted the latter part of chapter 2 may well now skip to chapter 6 for which all necessary preparation has been made. Sections in chapter 5 can be picked up as they are needed. If it is analysis of variance that is most interesting, skip to chapters 10 and 11.

4.4—Size of sample. In the formula for t , given near the end of section 4.2, division of both numerator and denominator by $n - 1$ gives

$$t = \bar{x}\sqrt{n/(2s^2)},$$

where s^2 is the pooled variance. If this equation is solved for n , then

$$n = 2t^2s^2/\bar{x}^2 = 2t^2C^2,$$

where C is the *relative* standard deviation or coefficient of variation (section 2.11). With the usual assumptions (section 2.14) one may substitute for t its 5% (or 1%) level then calculate the number of observations required in each group to produce this value of t . For discussion of changes in degrees of freedom see (3).

As an example, let us estimate the number of chicks (table 4.1) in each group necessary to make $t = 2.086$ provided C^2 does not turn out to be larger than $811/(41)^2 = 0.482$:

$$n = 2(2.086)^2(0.482) = 5 \text{ chicks per lot}$$

The nature of the uncertainty in making such estimates is emphasized by applying the formula to the data of the uniformity trial in example 4.1. If we did not know that these two samples were drawn from the same population we might estimate

$$n = 2(2.101)^2(14.44)/9 = 15 \text{ pigs}$$

as the number in each lot ample to show a significant difference. However, should we then draw the larger samples, we would likely get values of $\bar{x}$ and s^2 (known to be estimates of zero and 25 respectively) such that t would again be moderate in size, and the same would be true for any size of sample taken. But we might be misled. A hazard in applying this method is that we might happen to get one of those one-in-twenty samples with $t > 2$, then decide that the population difference is not zero, though it really is. Estimation of sample size with consequent repetition of the experiment is a common and useful practice, but it must be done with the canny caution necessary in all inquiries about the unknown.

EXAMPLE 4.10—The two lots of kernels in example 4.6 had $\bar{x} = 5$, $s^2 = 98.7$ and $d.f. = 28$. Estimate the number of kernels in each lot necessary to determine significance. Ans. 34 kernels.

EXAMPLE 4.11—If you were designing an experiment to evaluate decrease in concentration of nectar during flight, and if you considered the data of example 4.3 pertinent, what number of bees per lot might you try? Ans. 2 bees, according to the formula, but look at $t_{.05}$ for $d.f. = 2$.

EXAMPLE 4.12—Compare the formulas of sections 2.14 and 4.4. What theorem would enable you to infer either from the other.

4.5—Groups with different numbers of individuals. There is no necessity that the two groups be of the same size. In much experimentation it is inconvenient or impossible to provide equal numbers of individuals. Two lots of chicks, for example, from two batches of eggs differently treated would nearly always differ in numbers of birds hatched. Furthermore, accidents and death usually befall the lots unequally. This may not change the statistical theory, and causes only a slight alteration in the method of making the comparison.

As an illustration, consider the gains in weight of two lots of rats fed different levels of protein, table 4.4. If n_1 and n_2 are the numbers of

TABLE 4.4

GAINS IN WEIGHT OF TWO LOTS OF FEMALE RATS ON DIFFERENT DIETS. THE GAINS OCCURRED BETWEEN 28 AND 84 DAYS OF AGE

Diet	Number of Rats	Gains in Grams
High protein	12	134, 146, 104, 119, 124, 161, 107, 83, 113, 129, 97, 123
Low protein	7	70, 118, 101, 85, 107, 132, 94

observations in the two groups, the corresponding degrees of freedom are $n_1 - 1$ and $n_2 - 1$, a total of $n_1 + n_2 - 2$. The variances of the two group means are now different, s^2/n_1 and s^2/n_2 , while the variance of the difference of the means, the sum of the two variances, is

$$s^2/n_1 + s^2/n_2 = s^2(1/n_1 + 1/n_2) = s^2 \frac{n_1 + n_2}{n_1 n_2}$$

so that finally,

$$s_{\bar{x}} = \sqrt{s^2 \frac{n_1 + n_2}{n_1 n_2}} = s \sqrt{\frac{n_1 + n_2}{n_1 n_2}}$$

If $n_1 = n_2$, $s_{\bar{x}}$ becomes, as it should, the same as that in table 4.3. The necessary computations are laid out in table 4.5.

TABLE 4.5

SUMMARY OF STATISTICS FOR COMPARISON OF TWO GROUPS OF DIFFERENT SIZES. RAT DATA OF TABLE 4.4

Diet	Number of Rats	Degrees of Freedom	Mean Gain (grams)	Sum of Squares
High	12	11	120	5,032
Low	7	6	101	2,552
		Sum = 17	Difference = $\bar{x} = 19$	Sum = $Sx^2 = 7,584$

$$\text{Pooled variance} = s^2 = 7,584/17 = 446.12$$

$$s_{\bar{x}} = \sqrt{s^2(n_1 + n_2)/n_1 n_2} = \sqrt{446.12(12 + 7)/(12)(7)} = 10.04 \text{ grams}$$

$$t = 19/10.04 = 1.89, df. = 17, P = 0.08.$$

Since $P > 0.05$, the null hypothesis need not be rejected. So far as this small sampling is competent, the levels of protein fed may not differentiate rat gains into two populations.

The process of testing outlined above for groups of different sizes may be condensed into this formula:

$$t = \bar{x} \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{(n_1 + n_2) Sx^2}},$$

$\bar{x}$, as before, being the difference between the group means and Sx^2 the pooled sum of squares. Applying this formula to the rat gains: $\bar{x} = 120 - 101 = 19$ grams, $n_1 n_2 = (12)(7) = 84$, $n_1 + n_2 = 19$, $n_1 + n_2 - 2 = 17$, and $Sx^2 = 5,032 + 2,552 = 7,584$. Upon substituting these values in the formula, $t = 1.89$ as in table 4.5.

EXAMPLE 4.13—The following are the rates of diffusion of carbon dioxide through two soils of different porosity (9). Through a fine soil (F): 20, 31, 18, 23, 23, 28, 23, 26, 27, 26, 12, 17, 25; through a coarse soil (C): 19, 30, 32, 28, 15, 26, 35, 18, 25, 27, 35, 34. Show that pooled $s^2 = 35.83$, $s_2 = 2.40$, $d.f. = 23$, and $t = 1.67$. The difference, therefore, is not significant.

EXAMPLE 4.14—The total nitrogen content of the blood plasma of normal albino rats was measured at 37 and 180 days of age (11). The results are expressed as grams per 100 cc. of plasma. At age 37 days, 9 rats had 0.98, 0.83, 0.99, 0.86, 0.90, 0.81, 0.94, 0.92, and 0.87; at age 180 days, 8 rats had 1.20, 1.18, 1.33, 1.21, 1.20, 1.07, 1.13, and 1.12 grams per 100 cc. Compute $t = 8.2$.

EXAMPLE 4.15—Pearson and Catchpole (8) reported the means and standard errors of inorganic phosphorus per 100 ml. of blood serum as determined from 26 Percherons and 18 Shetlands, 3.29 ± 0.27 and 3.96 ± 0.40 mg. To test the significance of the mean difference, 0.67 mg., compute the pooled sum of squares, 96.345 (see example 4.8); $s_2 = 0.464$ and $t = 1.44$. The difference is not significant, but may be easily a consequence of sampling variation.

EXAMPLE 4.16—It is reported that the mean percentage of bacon from the carcasses of 15 Poland China sows was 6.59% with standard error, 0.300%. If you assume normal distribution, you can estimate the lowest and highest percentages in the sample as 4.6% and 8.6%. Would you have any reason to suspect the hypothesis of normality?

4.6—Null hypotheses and the validity of their tests. Occasionally there is reason to believe that the variances of the two treated populations are not the same. If they are not, the t -tests in tables 4.2 and 4.4 are nevertheless exact, but the null hypothesis being tested is this: *the two samples (groups) are randomly drawn from the same normal population; or in other words, the normal populations sampled have the same means and the same variances.* In this situation, a significant value of t may be attributed to a difference of means or a difference of variances or both.

Interest in an experiment may be enhanced by evidence that two treatments produce different variances. For instance, the following statistics were calculated from the data on nectar concentration in the honey bee, example 4.3:

Concentration of Syrup	Mean Decrease	Standard Deviation	Relative Standard Deviation, C
20%	0.5	0.163	33%
65%	1.9	0.767	40%

Not only are the population variances undoubtedly different (section 10.13 provides a test) but even the coefficient of variation is greater in the group with the larger mean. This kind of information may be more valuable than that about the difference between the means.

Despite this possibility, the fact is that the majority of investigators wish to know if their treatments affect the *means* of their population, *irrespective of variance*. They wish to test the hypothesis, $m_1 - m_2 = m = 0$, whether or not the samples are from populations with identical variance. If the hypothesis of equal variances is true, then the t -tests given above are exact tests of the hypothesis, $m = 0$. If there is reason to suspect that the variances are not equal, what shall be done? The answer is somewhat complicated, and will be given in three parts.

1. There are two classes of investigators who use tests of significance. Most of the readers of this book are among those who look upon such a test as an integral part of the evidence upon which decisions are based, but often as only a small part. For the moment, these may be called *casual* testers. Then there are *routine* testers who, planning a series of trials, adopt the invariable rule: reject the hypothesis if the probability is less than 0.05 (or 0.01). For these latter, it is essential to have the probabilities evaluated as precisely as possible, otherwise the proportion of rejections may be biased. They will be especially interested in references (1), (6), and (12), which are discussions of modifications in the testing appropriate for heterogeneous variance.

2. For the purpose of testing the means of two groups in which $n_1 = n_2$, the casual tester may, with one exception, ignore any difference in the variances of the groups. *Exception: enter t -table with $d.f. = n - 1$ instead of $2(n - 1)$.* With this exception, values of probability near the 0.05 level, got by the procedure of table 4.3, are likely affected by unequal variances no more than a point or two in the third decimal place. Decisions about the means, based on all the evidence, will be unchanged by heterogeneity of variance if $n_1 = n_2$.

3. If n_1 and n_2 , along with the corresponding variances, are greatly different, the casual tester may wish to take advantage of an easily applied approximate method proposed by Cochran and Cox (3). Since the procedure is somewhat different from that of table 4.5, it is illustrated in tables 4.6 and 4.7. The purpose is to test the hypothesis, $m = 0$, with no hypothesis about the population variances. The sums of squares are

TABLE 4.6
DECREASE IN CONCENTRATION OF NECTAR IN HONEY SACK OF BEE BETWEEN FEEDER
AND HIVE. TWO DAYS IN SEPTEMBER, 1931. FROM O. W. PARK (7)

Day in September	Number of Bees	Decrease in Percentage Concentration
3	20	1.0, 0.9, 0.8, 1.0, 0.8, 1.1, 0.6, 0.4, 0.4, 0.7, 0.9, 0.5, 0.9, 0.3, 0.3, 0.8, 0.6, 0.4, 1.0, 0.6
12	10	2.5, 0.6, 1.1, 0.6, 1.8, 0.6, 1.2, 1.2, 0.4, 1.0

TABLE 4.7
APPROXIMATE METHOD OF TESTING THE HYPOTHESIS, $m = 0$, WITH
NO HYPOTHESIS ABOUT VARIANCE

Date	Number Bees	Degrees of Freedom	Mean Percentage Decrease	Sx^2	$s^2 = \frac{Sx^2}{df}$	$s_{\bar{x}}^2 = s^2/n$	$t_{.05}$
9-3	20	19	1.1	1.24	0.0653	0.00326	2.093
9-12	10	9	0.7	3.72	0.4133	0.04133	2.262
	30	28	$\bar{x} = 0.4$			0.04459	

$$s_{\bar{x}} = \sqrt{0.04459} = 0.211$$

$$\bar{x}/s_{\bar{x}} = 0.4/0.211 = 1.90$$

$$5\% \text{ level} = \frac{(0.00326)(2.093) + (0.04133)(2.262)}{0.00326 + 0.04133} = 2.25$$

not pooled but the variance of each mean is calculated separately. The variance of the mean difference is, as usual, the sum of these variances. But the familiar ratio, $\bar{x}/s_{\bar{x}}$, is not now distributed like t with $df. = 28$, so cannot be compared to $t_{.05} = 2.048$. Instead, an approximation to the 5% level of this ratio is computed as the weighted mean of the two t -values for $df. = 19$ and $df. = 9$, the weights being the two variances, $s_{\bar{x}}^2$. That is, the sum of the products of the two pairs of numbers in the last two columns is divided by the sum of the variances. Since the ratio, $\bar{x}/s_{\bar{x}} = 1.90$, is smaller than its 5% level, 2.25, the hypothesis, $m = 0$, is not likely rejected. On the contrary, if the ordinary t -test is used to test the hypothesis that the samples were drawn from the same normal population (section 4.5), then

$$t = 0.4 \sqrt{\frac{(20)(10)(28)}{(30)(4.96)}} = 2.45, df. = 28,$$

and this hypothesis would presumably be rejected. The variances of the populations are evidently different (for a test, see section 10.13), but the means may not be.

4.7—Individual vs. group comparisons. The decision as to whether an experiment should be set up for individual or group comparisons is one of considerable import. It may even determine the success or failure of the trial. While there is no universal formula for determining the preferable design, there are a number of relevant circumstances on which judgments may be founded. Some of these follow.

Pairing is often indicated clearly in the experimental material. Crop yields in successive seasons tend to be so different that no one would think of comparing the results of one treatment last year with those of a

second treatment this year. Both treatments must be applied every season with a consequent pairing of the yields. Much the same can be said of different localities in the same year. Again, in the many experiments involving two measurements on the same individuals the measurements are nearly always paired. An example is the weighing of animals in two successive months in order to determine gain in weight.

On the other hand, pairing is not practicable in a variety of experiments. Among Breneman's chicks (section 4.2) there was no basis for matching individuals, hence, they were randomly assigned to the treatments. When yearling steers are shipped in from the range for fattening, so little is known of their history that estimating outcome seems to be doubtful. Again, if circumstances dictate different numbers of individuals for the two treatments, pairing is clearly impossible.

Aside from these rather obvious situations the design often depends upon the investigator's information about his experimental material and its environment. He may know, for example, that potted plants on two greenhouse benches will grow alike if treated alike. He could then apply one treatment to the pots on one bench and a second treatment to the others, then compare the two treatments by the methods of this chapter. But more often he will know that growing conditions on the two benches, and even on adjacent portions of the same bench, are different. He will avoid the effects of this variation by placing side by side pairs of pots with the two treatments, then replicating the pairs. The appropriate statistical method will be that of chapter 2.

With increasingly precise experimentation in any field, resulting from more exact knowledge of the behavior of the experimental material, group comparisons are likely to be replaced by those of individuals. In feeding trials with growing pigs, animals of the same litter, sex and initial weight are paired because it has been learned that they are more likely to gain alike on the same ration than are two randomly assorted individuals. Hence, if from each pair one member, chosen at random, is given one feed while the other member receives the second, the differences in gain tend toward uniformity. This means that the experimental error, calculated as in chapter 2, is expected to be smaller than that in a group comparison where the pigs are allotted to the rations at random.

In such experiments the success of pairing depends upon knowledge of outcome. No advantage will be gained unless the experimenter can predict with some degree of success that the members of a pair have the same performing ability, this common ability being different from that of the members of other pairs.

A feature of the method of individual comparisons is that it enables the investigator to utilize all the knowledge he has about the behavior of his experimental material in the contemplated environment. In the absence of this knowledge of outcome, or of any pertinent relationship, group comparisons are more appropriate.

The beginner may learn to avoid pitfalls by observing the effects of applying an unwarranted statistical method to experimental data. Suppose, for example, the reduction of the paired virus data in section 2.13 had been mistakenly carried out by the method for group comparisons. The results would have been these:

Preparation of Virus	Number Half Leaves	Degrees of Freedom	Mean Number of Lesions	Sum of Squares
1	8	7	15	468
2	8	7	11	172
		Sum = 14	$\bar{x} = 4$	$Sx^2 = 640$

$$s^2 = 45.71, t = 4\sqrt{(8)(7)/640} = 1.18, P = 0.26, \text{approximately}$$

The investigator might have decided not to reject the null hypothesis, whereas he probably would have rejected it on the basis of the correctly computed statistics in table 2.2. This discussion of misapplication of statistical methods should be compared with that of section 3.10 in which the experimental design was ignored. The results are identical, though differently evaluated.

Equally surprising effects follow an arbitrary pairing of data from an experiment designed for group comparison. When you remember that this design would be adopted only if the investigator were unable to foresee any advantage in pairing his experimental individuals, you will appreciate the mistake of arranging the randomly assembled measurements in pairs after the experiment is done. Perhaps the best way to illustrate is by use of the first two samples of table 3.3. Since these are two groups randomly drawn from the same population, the proper summary is as follows:

Sample Number	Number of Observations	Degrees of Freedom	Mean	Sum of Squares
1	10	9	35.6	$9(169.8) = 1,528.2$
2	10	9	29.3	$9(151.6) = 1,364.4$
Degrees of freedom = 18			$\bar{x} = 6.3$	$Sx^2 = 2,892.6$

$$s^2 = 2,892.6/18 = 160.7, t = 6.3\sqrt{9(10)/2,892.6} = 1.11, P = 0.28$$

The sample difference, 6.3, and variance, 160.7, are unbiased estimates of the population $m = 0$ and $\sigma^2 = 100$. Furthermore, as you would expect in 95% of such comparisons, t is less than its 5% value. But

suppose someone had rearranged the items in the two samples, pairing the two smallest, the next smallest, and so on to the two largest as in the first part of table 4.8; or else pairing small with large and *vice versa* as in the second part of the table. While the means of the samples and of the

TABLE 4.8
TWO OF THE POSSIBLE METHODS OF PAIRING THE DATA IN RANDOM SAMPLES,
ILLUSTRATING THE ERRONEOUS CONCLUSIONS THAT MIGHT BE DRAWN

	Method I			Method II		
	Sample 1	Sample 2	Difference	Sample 1	Sample 2	Difference
	12	11	1	12	53	-41
	24	19	5	24	44	-20
	29	19	10	29	32	-3
	33	24	9	33	31	2
	34	30	4	34	30	4
	36	30	6	36	30	6
	39	31	8	39	24	15
	39	32	7	39	19	20
	53	44	9	53	19	34
	57	53	4	57	11	46
ΣX	356	293	63	356	293	63
$\bar{x}$	35.6	29.3	6.3	35.6	29.3	6.3
s^2			8.01			627.34
t			7.04			0.80

differences are unchanged, the variances of the two sets of differences form a notable contrast, neither being unbiased estimates of the variance of random differences, $\sigma^2 = 200$, in the population. In the first method, the large value of t would certainly lead to rejection of the hypothesis of zero population difference. In the second, t is an underestimate of the correct sample value, 1.11. It is clear from these illustrations that the statistical method fixed by the experimental design cannot be tampered with.

4.8—Statistics and the experiment. It will be increasingly evident that the statistician and the experimenter, who are usually the same person, must form a close cooperation to insure the success of an investigation. The experimenter specifies the conditions in which the trial is to be performed—materials and treatments, together with genetic and environmental circumstances, and the measurements that can be made. The statistician selects or invents a plan (experimental design) which will furnish unbiased and unconfounded estimates with adequate precision. The experimenter conducts the laboratory or field work, taking pains to

eliminate as nearly as possible all extraneous effects. If he is successful, the ensuing measurements will contain the information for which the experiment is set up. The statistician uses appropriate methods for extracting all the information wrought into the data. Finally, the experimenter interprets this information in the light of existing knowledge in his science.

It is sometimes not understood that statistical methods can bring out only that information which has been incorporated into the data by careful design and execution of the sampling. Elaborate statistics are no substitute for meticulous experimentation. Population inferences are futile if dependent on carelessly collected data. It is equally often overlooked that extensive and conscientiously done measurements may contain little worthwhile information if the experimental design is faulty. It is only by a combination of appropriate design, skillful conduct of the experiment and suitable statistical methods that the investigator is assured of reliable evidence upon which to base his decisions.

4.9—Summary. Emphasis has been laid on the place of design in the conduct of any sampling. If the conditions are little known, the group comparisons of this chapter may be used. As the investigator's knowledge of experimental material and environment is enlarged he may utilize designs, such as those in chapter 2 and others to follow, that tend to eliminate from experimental error extraneous sources of variation.

In this chapter have been described statistical methods for group comparisons suitable to groups of both equal and unequal sizes. Unbiased estimates of differences between means and their variances are provided, followed by tests of hypotheses and fiducial statements.

REFERENCES

1. W. U. BEHRENS. *Landwirtschaftliche Jahrbucher*, 68:807 (1929).
2. W. R. BRENNEMAN. Personal communication.
3. W. G. COCHRAN and GERTRUDE M. COX. *Experimental Design*. Mimeographed (1944).
4. C. C. CULBERTSON and W. E. HAMMOND. *Iowa Agricultural Experiment Station Animal Husbandry Leaflet* 144 (1933).
5. HAROLD L. DEAN and R. H. WALKER. *Journal of the American Society of Agronomy*, 27:433 (1935).
6. R. A. FISHER. *Annals of Eugenics*, 11:141 (1941).
7. O. W. PARK. *Iowa Agricultural Experiment Station Research Bulletin* 151 (1932).
8. P. B. PEARSON and H. R. CATCHPOLE. *The American Journal of Physiology*, 115:90 (1936).
9. F. B. SMITH and P. E. BROWN. *Soil Science*, 35:413 (1933).
10. STEWART N. SMITH. *Journal of the American Society of Agronomy*, 26:792 (1934).
11. PEARL P. SWANSON and ARTHUR H. SMITH. *The Journal of Biological Chemistry*, 97:745 (1932).
12. P. V. SUKHATME. *Sankhya*, 4:39 (1938).

Chapter 5

SHORT CUTS AND APPROXIMATIONS

5.1—Standard method for computing standard deviation. When you were learning to compute standard deviations you were advised to carry no more digits in the deviations than were contained in the observed values. This led to small inaccuracies in the results, not serious in the foregoing computations. However, as the extent and complexities of the calculations increase, it becomes necessary to make them more precise. This is done in a very simple manner, by introducing a correction of just the right amount to compensate for the neglected digits. The method is now to be explained.

For illustration, take the birth weights of 20 guinea pigs from example 2.13. They are listed in the lefthand column of table 5.1. It is unnecessary to bother about computing the mean; just glance over the set of items and choose an *arbitrary origin*, G , near the middle of the array. Since 30 grams is near the mean, it will be chosen now for the first calculation. However, it will soon appear that 29 grams, or 40 grams, or any other arbitrarily selected number leads to exactly the same result after the proper *corrections for origin* are made.

The sum of the deviations from 30 is not zero, proving that 30 grams is not the mean of the set of birth weights. But the mean of these deviations, 0.25 gram, is just the necessary *correction for mean*, the amount that must be added to the arbitrarily chosen origin to arrive at the correct value of the mean, 30.25 grams. In example 2.13, the mean was calculated in the usual way by adding the column of weights and dividing by 20. By the short-cut method, the addition of the weights has been avoided. Perhaps a better attitude to take is that a check on the computations is made available. If the mean is calculated by two methods instead of one, the identity of the results proves the arithmetic.

In the last column of the table are the squares of the deviations from 30, together with their sum, 247. What we require is $\sum x^2$, the sum of squares of deviations from mean, usually abbreviated to *sum of squares*. The *correction for sum of squares* is made by the simple deduction of $[S(X - G)]^2/n = 5^2/20 = 1.25$. The sum of squares, then, is $247 - 1.25 = 245.75$. This may be verified by computing the 20 deviations from mean, $30 - 30.25 = -0.25$, etc., and adding the resulting squares. A less tedious exercise will be provided in example 5.4. The mean square (variance) and standard deviation are now calculated in the usual style.

Some people, unaccustomed to computation, confuse two of the quantities required in the foregoing calculations. First, there is the sum of the

TABLE 5.1
COMPUTATION OF STANDARD DEVIATION BY THE USE OF AN ARBITRARILY
CHOSEN ORIGIN OF MEASUREMENT
Birth weights of 20 guinea pigs in litters of two

Birth Weights (grams)	Deviations From 30 Grams		Squares of Deviations
X	$X - G$		$(X - G)^2$
30	0		0
30	0		0
26	-4		16
32		2	4
30	0		0
23	-7		49
29	-1		1
31		1	1
36		6	36
30	0		0
25	-5		25
34		4	16
32		2	4
29	-1		1
28	-2		4
27	-3		9
38		8	64
31		1	1
34		4	16
30	0		0
Total	-23	28	247

$$\begin{aligned}
 S(X - G) &= 28 - 23 = 5 \\
 [S(X - G)]/n &= 5/20 = 0.25 \\
 \bar{x} &= G + [S(X - G)]/n \\
 &= 30 + 0.25 \\
 &= 30.25 \text{ grams}
 \end{aligned}$$

$$\begin{aligned}
 S(X - G)^2 &= 247 \\
 [S(X - G)^2]/n &= 5^2/20 = 1.25 \\
 Sx^2 &= 245.75 \\
 s^2 &= 245.75/19 = 12.93 \\
 s &= 3.60 \text{ grams}
 \end{aligned}$$

squares of deviations from G , $S(X - G)^2 = 247$. Second, there is the square of the sum of the same deviations, $[S(X - G)]^2 = 25$. In the first the squaring is done before the addition, while in the second the order of the operations is reversed. The objectives and results are so different that it is only necessary to direct attention to the speciousness of the analogy.

EXAMPLE 5.1—Carry through the computations in table 5.1 with 25 grams as the arbitrary origin. The deviations will be 5, 5, 1, 7, 5, -2, etc. Verify these results: (i) $S(X - 25) = 105$ grams; (ii) $S(X - 25)^2 = 797$; (iii) correction for mean = 5.25 grams; (iv) correction for sum of squares = 551.25. The mean and sum of squares will be 30.25 grams and 245.75 as before.

EXAMPLE 5.2—Make the computations in table 5.1 with 35 grams as the origin. The deviations are -5, -5, -9, -3, etc. The correction for mean is $-95/20 = -4.75$, so that the mean is $35 + (-4.75) = 35 - 4.75 = 30.25$ grams as before. The addition is performed algebraically as usual. The correction for sum of squares is $(-95)^2/20 = 451.25$, a positive number. This correction, then, is deducted from $S(X - 35)^2$.

EXAMPLE 5.3—In table 5.1, if you use the arbitrary origin 23, or more conveniently 20, you will have no negative deviations. Is there any advantage in computing with deviations which are all positive?

EXAMPLE 5.4—Here is an easy set for practice: 1, 4, 4, 7, 7, 8, 10, 10, 10, 12, 15. Verify the fact that the mean is 8 and the standard deviation, 4. Use various arbitrary origins, continuing until the process of calculation can be carried through without hesitancy. Try $G = 0$.

EXAMPLE 5.5—In section 3.5 it was suggested that you compute s in each of the samples you had drawn. You can now get the exact sample values by making the corrections in $S(X - G)^2$.

EXAMPLE 5.6—What evidence do you have that this theorem is true? The sum of squares of deviations from mean is less than the sum of the squares of the deviations from any other origin.

EXAMPLE 5.7—Might the correction for sum of squares be calculated just as easily by the quantity, $[S(X - G)]\bar{x}$? In words, might the phrase, "sum times mean," be substituted for "sum square over number"? Which do you prefer?

EXAMPLE 5.8—Those who enjoy algebraic manipulations will wish to derive the formula, $Sx^2 = S(X - G)^2 - [S(X - G)]^2/n$. Start with the relation, $x = X - \bar{x} = (X - G) - (\bar{x} - G) = (X - G) - S(X - G)/n$. Either transpose the term, $S(X - G)/n$, then square and sum, or square and sum the equation as it stands.

5.2—Special methods for use with calculating machines. The methods which have been described were devised to confine the computations, so far as possible, to small integers. If a calculating machine is available, large integers can be manipulated almost as easily as small ones. Other objectives, therefore, may be set up. One is the avoidance of negative numbers. Another is the elimination of most of the subtractions. Both of these objectives are easily attained, the resulting method being especially adapted to machine calculation. Here it is.

Returning to the weights in the first column of table 5.1, their sum, $SX = 605$, and the sum of their squares, $SX^2 = 18,547$, are readily calculated with any machine. The details of the operations differ with the style of machine used. Usually the two sums are run up simultaneously, the appropriate method being described in the manual of instructions issued by each manufacturer. With any machine it is unnecessary to copy down the squares of the individual weights; the sum is accumulated in the machine as the operation proceeds. At the end, it is only necessary to copy the results from the dials into a table like 5.2. The mean, of

TABLE 5.2
METHOD OF COMPUTING THE MEAN AND SUM OF SQUARES IF A
CALCULATING MACHINE IS USED
Birth weights of 20 guinea pigs as recorded in table 5.1

$n = 20$	$SX^2 = 18,547$
$SX = 605$	$(SX)^2/n = 18,301.25$
$\bar{x} = 30.25$ grams	$Sx^2 = 245.75$

course, is $(SX)/n = 605/20 = 30.25$ grams. The correction for sum of squares is $(SX)^2/n = (605)^2/20 = 18,301.25$. Deducting this from SX^2 , the corrected sum of squares is 245.75 as before.

This method of machine calculation may be thought of as a special case of the general method described in section 5.1. Now, $G = 0$, the arbitrary origin being the same as the zero point on the scale of weights. The deviations, $X - G = X - 0 = X$, are the same as the observed values. The correction necessary to obtain the sum of squares is $[S(X - G)]^2/n = [S(X - 0)]^2/n = (SX)^2/n$. Not only is the machine method a timesaver, but, obviating as it does the subtractions associated with deviations, it affords less chances for error.

EXAMPLE 5.9—Using the machine method, verify $\bar{x}$ and s in examples 2.6 and 2.14.

EXAMPLE 5.10—Using the machine method, verify $\bar{x}$ and s in example 2.24.

EXAMPLE 5.11—If a machine is used for calculating the standard deviation of the differences in table 4.2, what change in the operation would be required by the negative differences?

EXAMPLE 5.12—Might the correction for SX^2 in table 5.2 be calculated just as easily from the phrase "sum times mean" (example 5.7)? In symbols, is $(SX)^2/n = (SX)\bar{x}$?

EXAMPLE 5.13—It is easy to derive the formulas, $Sx^2 = SX^2 - (SX)^2/n = SX^2 - \bar{x}SX = SX^2 - n\bar{x}^2$. Start with the relation, $x + \bar{x} = X$.

EXAMPLE 5.14—In example 2.6 were recorded the differences in yield on 16 pairs of plots in 1933 and 1934. In the three preceding years there were 25 other pairs, with differences 9.0, 1.4, 4.1, 0.0, 6.3, 4.6, 1.1, -1.2, 2.0, 2.7, -1.6, 6.3, 4.5, 9.1, 0.4, -0.2, 2.8, 8.6, 4.1, 6.0, 8.1, 10.6, 6.2, 8.7, 8.4 bushels per acre. Compute the mean of the 41 differences, 4.18, and the standard error, 0.615 bushels per acre. Is the difference significant?

5.3—Deviations as code numbers. Even when a calculating machine is available, it is occasionally desirable to use deviations from some easily subtracted arbitrary origin. As an example, consider the weights of 10 tomatoes: 206, 217, 224, 227, 228, 231, 236, 241, 245, 258 grams. It is easy to subtract 200 grams from each item without rewriting the set. The machine is used to run up $SX = 313$ and $SX^2 = 11,741$ for the deviations, 6, 17, . . . 58. The mean, then, is $200 + 313/10 = 231.3$ grams, the sum of squares being $11,741 - (313)^2/10 = 1944.1$. There is a worthwhile saving in the size of the numbers, with little chance of error in making the subtractions.

The deviations from 200 are sometimes called *code numbers*. In the computations they are used instead of the observed values in order to facilitate the operations. If deviations were the only kind of code numbers, there would evidently be no advantage in having a second name for them. A rather obvious application of another kind of code number is shown in table 5.3. For present purposes, the hydrogen ion concentrations (pH) are ignored. The readings in the second column are coded by: (i) multiplication by 10,000, and (ii) subtraction of 200. The computations of $\bar{x}$

and s are done with code numbers. The resulting mean must first be increased by 200 then divided by 10,000, these inverse operations* being applied in reverse order to the coding operations. You will notice that s is not affected by the subtraction of 200; but it, like the mean, must be

TABLE 5.3
THE USE OF CODE NUMBERS IN CALCULATING MEAN AND STANDARD DEVIATION
Cyanogenetic nitrogen recovered from peach stems (2)

Initial pH	Grams per 50 Grams of Tissue X'	Code Numbers $X = 10,000X' - 200$
5.1	0.0238	38
5.0	0.0238	38
5.3	0.0241	41
5.3	0.0241	41
5.7	0.0250	50
5.7	0.0247	47
6.0	0.0241	41
6.1	0.0238	38
6.6	0.0226	26
6.5	0.0232	32

$n = 10$	$SX^2 = 15,784$	$s = 6.81$ code numbers
$SX = 392$	$(SX)^2/n = 15,366.4$	$= 0.000681$ gram
$\bar{x} = 39.2$ code numbers	$Sx^2 = 417.6$	
$= 0.02392$ gram	$s^2 = 46.4$	

compensated for the coding operation of multiplication by 10,000. Coding by multiplication or division may be thought of as a change in the unit measurement. If it is easier to carry on the calculations with a unit of one ten-thousandth gram, little chance of error is introduced by the change. At the end, the original unit is recovered by a mere shift of the decimal point.

5.4—True coding. The really practical use of code numbers arises from a custom general among researchers. Both from necessary precaution and from convenience they are prone to collect their data in finer units than are suitable to statistical treatment. A good example appears in the report of a survey of 145 Iowa farm families with the object of learning about their habits of food consumption (6). Naturally, the record of the annual consumption of meat was in pound units, that of the first 20 families being 726, 296, 928, 668, 287, 1206, 517, 1638, 2414, 610, 494, 2489, 1198, 676, 1302, 440, 1247, 1053, 1029, 218 pounds. The range is $2489 - 218 = 2271$ units. In most statistical investigations a range somewhere between 20 and 40 units is ample. It turns out, then, that the hundredweight is a more suitable computing unit for the set above, the appropriate code numbers being derived from dividing each item by 100,

* Addition and subtraction are *inverse* operations; so are multiplication and division.

then *rounding*: 7, 3, 9, 7, 3, 12, 5, 16, 24, 6, 5, 25, 12, 7, 13, 4, 12, 11, 10, 2 cwt. The statistics of this coded sample are, $\bar{x} = 9.65$, $s = 6.36$ cwt. Of course, these must be multiplied by 100 to recapture the original unit, so that $\bar{x} = 965$ pounds and $s = 636$ pounds.

In rounding, numbers like 13.51 are increased to 14, while such as 13.49 are lowered to 13. The dividing line being 13.50, what shall be done if this particular number appears? In order to insure increases and decreases in approximately equal proportions a good rule is to round to the nearest even number. Thus, 13.50 and 14.50 would each be rounded to 14.

You will immediately wish to know whether any valuable information has been lost owing to the rounding of the code numbers. The statistics of the original sample are $\bar{x} = 972$, $s = 637$ pounds. Is the discrepancy disturbing? Consider the standard error of the mean, $s_{\bar{x}} = 637/\sqrt{20} = 142$ pounds.

EXAMPLE 5.15—The dry weights of clippings from five bluegrass plants (4) for the week ending July 7 were 1.2, 0.9, 1.0, 1.4, 0.9 grams. Before computing $\bar{x} = 1.08$ and $s = 0.217$ grams, imagine each item multiplied by 10; that is, ignore the decimal points. The correct results are got from dividing the coded statistics, $\bar{x}$ and s , by 10.

EXAMPLE 5.16—Do you have a name for such quantities as SX^2 and $S(X - G)^2$? You can't use "sum of squares" because that is reserved for Sx^2 . You might say "sum of squares of the observations" and "sum of squares of deviations from G ."

Here are three examples for those interested in the algebra of the preceding and following sections.

EXAMPLE 5.17—If $Y = A + X$, A being constant, prove that $\bar{y} = A + \bar{x}$ and $Sy^2 = Sx^2$.

EXAMPLE 5.18—If $Y = bX$ (b is another constant), prove that $\bar{y} = b\bar{x}$ and $Sy^2 = b^2Sx^2$. You won't overlook the fact that b may be a fraction.

EXAMPLE 5.19—If $Y = A + bX$, prove that $\bar{y} = A + b\bar{x}$ and $Sy^2 = b^2Sx^2$.

5.5—Rules and precautions about code numbers. The objects of coding are to promote accuracy and save labor. Be sure that these advantages are great enough to more than counterbalance the time and risk of error involved in coding.

Every item in a set must be treated alike. If you were to add one number to some of the items and a different number to the others, the results would be meaningless.

The mean is affected by every coding operation. As an example, if items are coded by first subtracting 100 then dividing by 10, the coded mean must be first multiplied by 10 then increased by 100, the inverse operations being applied in reverse order.

The standard deviation is affected by only multiplication and division. It should be clear that addition and subtraction do not affect measures of variation such as range or s ; such coding merely shifts the origin of measurement without contracting or expanding the unit. On the other hand, multiplication and division change the unit of measurement and must be compensated by applying the inverse operations to the coded statistics.

The precision of the final result is not affected by addition, subtraction, multiplication or division, but is influenced by dropping off digits and rounding.

The inference was made in section 5.4 that a range somewhere between 20 and 40 units is desirable. The reference was to code numbers, since they do not ordinarily involve decimals. A more general statement is that the range when expressed in significant digits should lie somewhere between 20 and 40. Significant digits (the term has no technical kinship with that employed in sections 1.13, 3.9, and 4.2) are independent of the position of the decimal point. The significant digits are the same in the numbers 647, 0.0647 and 6.47. If the range in one set is $0.067 - 0.031 = 0.036$ gram, and in another, $23.1 - 19.5 = 3.6$ pounds, the ranges in significant digits are the same, 36. When considering the number of digits to retain for computation, the decimal point may be ignored until it is time to make a record of the plan selected.

EXAMPLE 5.20—In example 6.33 are given percentages of water in apple twigs and corresponding measurements of specific heat. If you have a calculating machine available, compute $\bar{x}$ and s for each character. Note especially that you cannot round off the last digit in either column. If you were to do so, the range of rounded numbers in column 2 would be only $65 - 49 = 16$, which is less than desirable. Of course, you may omit the decimal point, then subtract 400 from each percentage if you wish. With a machine, however, the saving in time is questionable unless you are accurate enough to carry on the coding while operating the machine and without rewriting the column. Try it.

5.6—Significant digits. “How many decimals should be carried” is a question often asked. In that form there is no definite answer. When the question is put in the form, “How many significant digits should be carried?” then answers are possible though various. The distinction, pointed out in the preceding section, is made clearer by the following example. A weighing is made on a balance read directly to fifths of a gram, but estimated to tenths. The weight is 25.3 grams. But this result might be recorded as either 25,300 mg. or 0.0253 kg. In any of the three forms the significant digits are 253, quite unaffected by the unit of measurement. Even when the question above is limited to significant digits, however, there is an amazing array of considerations affecting the number of figures carried. It seems quite impossible to reduce them to rules. Throughout this text we have endeavored to display good form in handling numbers. At various points we have commented on unusual complications. In this section, we shall attempt little more than to suggest some trains of thought and some experiments in calculation.

If 25.3 g. is the dry weight of some forage plant grown in a pot, it may be necessary to convert the figure into pounds per acre. The conversion factor might be 94.327 pounds per acre per gram. How many significant digits should be carried in the product? Remembering that 25.3 grams really is an estimate of some weight between 25.25 and 25.35, the number of pounds per acre may be anywhere between

$$(25.25)(94.327) = 2,381.8 \text{ lbs. per acre}$$

and $(25.35)(94.327) = 2,391.2 \text{ lbs. per acre}$

The mean of these two products is equal to $(25.3)(94.327) = 2,386.5$. Three facts are clear: (i) even the third significant figure may not be right, but (ii) it takes four significant figures 2,386 to estimate the product satisfactorily, and (iii) the factor with the fewer significant figures limits the accuracy of the product—the weakest-link-in-the-chain idea. It is easy to verify the fact that the product of two three-figure measurements is even less accurate than the result considered above, but that the fourth figure still gives some information about the average value. A good rule in multiplication, then, is to carry one more significant digit than those in the shorter of the two factors.

You will find by a little experimentation that the facts about quotients are much the same as those about products. But consider the combination of subtraction and division illustrated in a report of the percentage decrease in amino nitrogen in a cabbage plant extract during distillation with calcium oxide (8). The number of milligrams before distillation was 14.3; after, 10.8, the percentage decrease being 24.5. At the limits of approximation, the decrease was between

$$\begin{array}{rcl} & 14.35 - 10.75 = 3.60 \text{ mg.} \\ \text{and} & 14.25 - 10.85 = 3.40 \text{ mg.} \end{array}$$

The percentage decrease was, therefore, between

$$\begin{array}{rcl} & (3.60)/14.35 = 25.1\% \\ \text{and} & (3.40)/14.25 = 23.9\% \end{array}$$

Thus, inaccuracy has crept into the second significant figure, all beyond the third being meaningless.

In computing the sum of a number of items similar to 25.3 grams, since there are likely to be about as many weights too small as too large, there is a tendency for the excesses to cancel the defects. The mean, therefore, can be carried safely one or two significant figures farther than the items.

Rules could be devised for the computations discussed above but they would have to be discarded promptly when a whole series of calculations is to be performed, as in the computation of the standard error of a mean. Continuous adherence to the rules would result in the introduction of exaggerations of inaccuracies. To be on the safe side, two or three extra figures must be carried, especially if calculating machines are available. The final results should be cut back to a reasonable number of significant figures.

What is a reasonable number? The errors of measurement considered above are relatively insignificant when compared with sampling variation. As an example, a mean, 55.957 mg., may be accurate to five significant digits in a large sample, and its standard error may be correctly computed as 1.754 mg. But since the 50% semi-fiducial interval is $(0.6745)(1.754) = 1.2$ mg., the report might well be limited to 56.0 ± 1.8 mg. so far as inferences about the population are concerned. However, that brings up

another complication. If we wish to reconstruct the normal sample from which these statistics were calculated, in the manner of example 4.8 for instance, it would be advantageous to have at least as many significant figures in s_2 as there are in $\bar{x}$, the actual number depending somewhat upon the size of the sample (since s_2 is to be multiplied by $\sqrt{n}$).

You begin to see why one can't follow any one set of specialized rules about significant digits. In simple products and ratios, it is easy to make the results compatible with the precision of the original measurements. If you are making a long sequence of calculations for the first time, carry along several more figures than you expect to retain in the final form. In preparing your report, limit all numbers to a size commensurate with their precision and contemplated utility.

EXAMPLE 5.21—In making the weighings of dressed swine carcasses, and especially of live weights, accuracy beyond whole pounds is seldom attained. What would you think if you saw the report of a dressing percentage, 81.764%?

EXAMPLE 5.22—Suppose, in calculating the sum of squares of 15 two-figure items, $\sum X = 545$, so that $\bar{x} = 36.33$. The correction for sum of squares is, then, $(545)^2/15 = 19,801.67$. If you had used the formula of example 5.7, you would have got $(545)(36.33) = 19,799.85$. Hence, if you say "sum times mean" is the correction term, you must carry the mean to more significant figures than you expect to report. If you wish to get the foregoing correction term accurate to two decimal places, how many significant figures will be required in the mean? Note that this is not a question of sampling variation, but is merely a matter of correct calculation. The rounded off figures in the mean, 0.00333 . . . , are magnified 545 times by multiplication, and $(545)(0.00333 \dots)$ is 1.81. If you wish to use the formula "sum times mean," and it is very convenient at times, you must be sure to carry the mean to an adequate number of significant figures.

5.7—The distribution of the ratio, range/s. Turning from methods of computation, we introduce a topic of great utility in the common sense understanding of statistics. As has been emphasized, the two measures of variation, range and standard deviation, are not unrelated. In example 3.1 it was suggested that you calculate the ratio, *range/s*, for each of your samples of 10. Table 5.4 is the frequency distribution of 500 such ratios in samples similar to yours. If you make the histogram you will be struck by the asymmetry of the distribution as well as by the magnitude of the range. You can easily find where your ratios belong in the table—which are near the average and which among the extremes?

TABLE 5.4
FREQUENCY DISTRIBUTION OF *range/s* IN 500 SAMPLES OF 10
Mean = 3.18

Class mark	1.95	2.05	2.15	2.25	2.35	2.45	2.55	2.65	2.75	2.85	2.95
Frequency	1	0	1	0	1	2	8	24	25	43	44

Class mark	3.05	3.15	3.25	3.35	3.45	3.55	3.65	3.75	3.85	3.95	4.05
Frequency	58	61	55	45	45	36	27	12	9	2	1

If you should investigate the corresponding distributions for samples of other sizes, you would find that the average value of the ratio, $range/s$, tends to increase with sample size. Why? As more items are drawn, there is an increasing likelihood of including some of the unusual extremes which are in the population. Both range and standard deviation tend to increase with sample size, but range goes faster.

Tippett (9) and E. S. Pearson (7) have examined the distribution of $range/\sigma$ in samples from normal populations. Tippett tabulated the mean values and Pearson, the 5% and 1% points for samples of many sizes: table 5.5 is taken from their reports. In practice, it is only $range/s$ that is known, but attention is called to the close agreement between the mean value of $range/s = 3.18$ in table 5.4 and the mean of $range/\sigma$, 3.08, in table 5.5. In the kinds of applications made in this text, the discrepancy is unimportant.

TABLE 5.5

MEAN VALUES OF THE RATIO, $RANGE/\sigma$, TOGETHER WITH VALUES THAT WILL BE EXCEEDED WITH INDICATED PROBABILITY IN RANDOM SAMPLES OF SIZE n DRAWN FROM THE NORMAL POPULATION

n	Probability			n	Probability		
	Mean	.05	.01		Mean	.05	.01
2	1.13	2.77	3.64	20	3.73	5.01	5.64
3	1.69	3.34	4.10	30	4.09	5.30	5.91
4	2.06	3.65	4.38	50	4.50	5.64	6.23
5	2.33	3.87	4.59	75	4.81	5.91	6.47
6	2.53	4.04	4.74	100	5.02	6.08	6.63
7	2.70	4.18	4.87	150	5.3	6.3	6.8
8	2.85	4.29	4.98	200	5.5	6.4	7.0
9	2.97	4.39	5.07	300	5.8	6.7	7.2
10	3.08	4.48	5.15	500	6.1	6.9	7.4
15	3.47	4.79	5.44	700	6.3	7.1	7.6

Although table 5.5 already has been made use of considerably, one additional example may be cited. Crampton (1) reported the mean area of rasher of 20 swine as 26.1 square inches, with standard deviation 2.8 square inches. Most of us have only a vague notion of the range in area of slices of bacon. In table 5.5, opposite $n = 20$, the ratio 3.73 is read. Therefore, average range = $(3.73)(2.8) = 10.4$ square inches. The largest and smallest rashers are estimated as $26.1 \pm (10.4/2)$; that is, 31.3 and 20.9 square inches. Actually, the extreme areas in the sample as reported were 32.8 and 21.3 square inches.

The table allows the estimate not only of the average range but also of the ranges that will be exceeded in only 5% and 1% of random samples from the normal population. Thus, it may be said that the chance of exceeding the range of rasher area $(5.01)(2.8) = 14$ square inches is only 5% if samples of 20 are randomly drawn from a normal population with $\sigma = 2.8$ square inches.

EXAMPLE 5.23—The mean yield of 207 plots of Kherson oats in 1916 (5) was 74.8 bushels per acre, the standard deviation, 6.2 bushels per acre. Assuming normal distribution, show that the estimated range is 34 bushels per acre. Actually, the range was from 56.7 to 92.8.

EXAMPLE 5.24—The range of percentage dry matter in 48 samples of Grimm alfalfa was 7.6%. One would estimate the standard deviation as 1.7%.

EXAMPLE 5.25—It was reported that the standard deviation of a sample of cuckoos' eggs was 0.964 mm., the range in length being from 19.07 to 25.02 mm. Would you estimate the sample size as about 600 eggs?

5.8—Common sense in computation. I have never found a magic for insuring accuracy in computations. Despite checks and double checks, errors may creep in. A correction term may have been forgotten or a decimal point misplaced. The final safeguard must be good judgment as to the credibility of the results.

Suppose someone has calculated the standard deviation of the heights of 200 10-year-old girls. His result is 1.2 inches. He wishes to know if this is reasonable. Looking in table 5.5, opposite $n = 200$, he takes out the ratio 5.5, from which he knows that, on the average, the range should be given by, $\text{range} = (5.5)(1.2) = 6.6$ inches. A glance at his records reveals the actual range from 45 to 58 inches. Certainly, he has cause to suspect his calculated standard deviation. Another approach is this. Start from the known range, $58 - 45 = 13$ inches. Substituting this range together with the ratio read from the table, the equation becomes, $13/s = 5.5$, from which estimated $s = 2.4$ inches. This is a warning that the calculated $s = 1.2$ inches is probably too small.

Most people have reasonably accurate knowledge about means and ranges in their experimental materials. Table 5.5 enables them to add a practical knowledge of standard deviations. It is sufficient to fix in mind four approximate ratios—3, 4, 5, and 6—corresponding to samples of 10, 25, 100, and 500 items. If the work is with rats, the investigator is familiar with the mean weight of males at weaning, 52 grams, and the range which is roughly $90 - 30 = 60$ grams. The standard deviation of 100 such animals would, therefore, approximate $60/5 = 12$ grams.

The same approximate ratios are valuable to a person reading of the researches of others. He may see the report that the mean number of stamens in 373 late flowers of *Ranunculus ficaria* was 18, with standard deviation 3.3. Assuming normal distribution, he estimates the range as between five and six times 3.3, say 17 or 18 stamens. He is thereby informed that the smallest number of stamens in any flower was about $18 - 18/2$, or 9 stamens. Of course, he would not be surprised to learn from some other source that the smallest number was actually 7, or perhaps 12 stamens. The point is that he can sketch in the outlines of the information possessed by the author of the article.

One can learn to judge of significance from surprisingly meager information. It may be reported that two lots, each consisting of 10 swine, fed different rations, produced back fats with iodine numbers averaging 68 and 73 respectively. If the reader of the report, from his own work in the

same field, had the information that iodine numbers range an average of 10 points in samples of 10, he approximates the standard deviation as 3.3 iodine numbers. Proceeding as in chapter 4, he estimates the mean difference as highly significant. Or, he reads that the mean height of 800 13-year-old girls exceeds that of an equal number of boys of the same age by 1.3 inches. He wishes to know if girls of this age are really the taller, or if the difference may be attributed to the vagaries of sampling. Knowing that stature, even at this age, has a standard deviation not greatly different from 2.5 inches, he estimates the standard error of the difference, $2.5\sqrt{2/800} = 0.125$ inch. It is clear that, at 13 years, taller girls is a population characteristic.

EXAMPLE 5.26—One thousand butterfat determinations were made on the milk from a herd of Holsteins, the lowest being 2.2% and the highest 5.3%. Would you estimate the mean and its standard error as 3.75 ± 0.015 percent?

EXAMPLE 5.27—Twenty swine were paired according to litter, sex, weight and condition. One member of each of the 10 pairs was fed ration A, the other B. The mean difference in gain at the end of 60 days was 5.75 pounds the range being from -7 to 20 pounds. Show that, according to the rough method of estimation which is available, the mean difference fails to reach the 5% level of significance.

EXAMPLE 5.28—An asparagus grower recorded the production of 25 pairs of staminate and pistillate asparagus plants during one season. He found that the difference, $M - F$, ranged from -10 grams to 90 grams. Estimate $\bar{x} = 40$ and $s = 25$ grams. Show that the difference is highly significant.

EXAMPLE 5.29—Twenty soybeans were planted on each of 100 plates and placed in a germinating compartment (3). Suppose the standard deviation of the number of beans germinating on each plate had been computed as 6 beans. Show that the calculation was in error, assuming a somewhat normal distribution of germinating seeds.

EXAMPLE 5.30—The standard deviation of the weights of 1,000 college freshmen men was calculated as 16 pounds. Is that a reasonable figure?

EXAMPLE 5.31—Sixty-four samples of peanuts were analyzed for nitrogen content, the mean and its standard error being $11.04 \pm 0.117\%$. Would you be surprised if told that one of the samples contained only 5% nitrogen?

5.9—Rates and percentages. The habit of expressing one measurement as a fraction of another is as common as the use of ratios in enumeration data (section 1.16). Such quotients are called *rates* if the numerator and denominator have different units of measurement, and *ratios* or *percentages* if the units are the same. The yield of a crop, a rate, is almost always reported in some such fashion as 30 bushels per acre, while the dressed weight of a carcass is, say, 80% of the live weight. These rates and ratios are summary numbers or averages. In using them the same precautions should be observed as those discussed in section 1.16. A distinction is this: there is no test for significance of the difference between two isolated quotients of measurement data. One can't test directly the significance of the difference between 30 and 25 bushels per acre, or between 3.5% and 4% milk. Only if there is available additional information concerning variation can the test be made. The method then may be that of either chapters 2 or 4 depending upon the experimental design.

One rate interesting to many is the rate of growth of an organism. Growth rates are sometimes expressed as changes in weight (or size, etc.) per unit time and sometimes as changes in weight *per unit time per unit weight already attained*. The latter is called *relative growth rate*. As an example, the growth rate of a pig between 60 and 120 days of age, kept under uniform conditions, may be rather constant at 1.4 pounds per day. During the same 60 days the relative growth rate is declining, starting (if the initial weight is 70 pounds) at some such amount as

$$\frac{1.4 \text{ pounds per day}}{70 \text{ pounds weight}} = 0.02 \text{ pound per day per pound,}$$

and ending at

$$\frac{1.4 \text{ pounds per day}}{70 + (60)(1.4) \text{ pounds weight}} = 0.009 \text{ pound per day per pound.}$$

But earlier in the life of the pig a different picture may be seen. The actual growth rate may increase from day to day in such manner that the relative growth rate is constant; that is, each pound may increase by the same amount every day. An example of uniform relative growth rate is given in section 14.2. One should distinguish carefully between these methods of expressing growth rate.

Another distinction that should always be clearly made is that between *amounts* of change (differences or increments) and *rates* of change. A pasture yielding 2 tons of hay per acre one year may produce 4 tons the next. Though the difference is 2 tons, the increase is sometimes expressed as 100%. As a rule, the expression of differences as percentages of a small base should be avoided.

EXAMPLE 5.32—Five dairy farmers delivered milk at a creamery in the following amounts:

Farmer	Pounds of Milk	Percentage Butterfat in Milk
A	150	4.75
B	500	3.50
C	350	4.00
D	400	3.00
E	100	5.00

Calculate the average percentage butterfat, 3.71.

EXAMPLE 5.33—Those who have some mathematics at their command will observe that if the natural logarithm of the weight, $\log_e w$, is plotted against time, t , the slope of the curve at any point,

$$\frac{d}{dt} \log_e w = \frac{1}{w} \frac{dw}{dt},$$

gives the relative growth rate. If common logarithms are used, the slope must be multiplied by $\log_{10} e = 2.3026$ to get the relative growth rate.

REFERENCES

1. E. W. CRAMPTON, Macdonald College, Quebec. Unpublished data.
2. O. W. DAVIDSON and J. W. SHIVE. *Plant Physiology*, 10:73 (1935).
3. T. I. EDWARDS. *Plant Physiology*, 9:8 (1934).
4. CARTER M. HARRISON. *Plant Physiology*, 9:94 (1934).
5. T. A. KIESSELBACH. *Nebraska Agricultural Experiment Station Research Bulletin* 13 (1918).
6. P. MABEL NELSON, ELIZABETH E. HOYT, LAURA McLAUGHLIN and ETHEL CESSNA MORGAN. *Iowa Agricultural Experiment Station Bulletin* 337 (1935).
7. E. S. PEARSON. *Biometrika*, 24;416 (1932).
8. NEIL W. STUART. *Plant Physiology*, 10:142 (1935).
9. L. H. C. TIPPETT. *Biometrika*, 17:386 (1925).

Chapter 6

LINEAR REGRESSION

6.1—Introduction. In foregoing chapters the problems considered have involved primarily only a single measurement on each individual. This measurement may be of vitamin content or weight in which the variate is continuous, or of color or opinion, the variate being discrete. The observations are compared either singly or in groups, but generally it has been one attribute or one kind of measurement per individual. Now attention is turned to inferences based on two or more observations on each member of the sample. For example, one may make more adequate judgments about the 120-day weight increase of a pig if he knows the related variates, initial weight, age, and feed eaten; increasing farm income during some periods may be explained partly by changes in the quantity of meat produced; or decisions about the comparative yielding abilities of varieties of corn may hinge on the numbers of plants in the several experimental plots. The widespread interest in correlated measurements arises from their numerous applications, only a few of which can be discussed in this chapter.

In the parlance of biological and mathematical statistics, Galton's descriptive term *regression* has largely replaced the technical word *function* used in mathematics. Growth curves and economic time series are referred to as *regressions of height (or income) on age (or time)*. Further comment on the origin of this usage will be found in section 6.13.

6.2—Regression of height on age. Twenty soybeans were planted in separate pots, and these were grouped on the greenhouse bench in 4 blocks of 5 pots each. A week after the plants emerged measurement was made of the heights of 4 of them, one chosen at random from each of

TABLE 6.1
HEIGHT (CENTIMETERS) OF GROUPS OF SOYBEAN PLANTS, ONE GROUP MEASURED IN
EACH OF 5 SUCCESSIVE WEEKS AFTER EMERGENCE

Week After Emergence	Block Number				Average Height
	1	2	3	4	
1	4	3	6	7	5
2	18	19	18	13	17
3	26	25	24	21	24
4	38	35	28	31	33
5	44	43	39	38	41

the blocks. Measurements on similar groups of four were made in each succeeding week, no plant being taken twice. The heights of the 20 plants are recorded in table 6.1, and the pairs of values of the two variates, age and height, are plotted in figure 6.1.

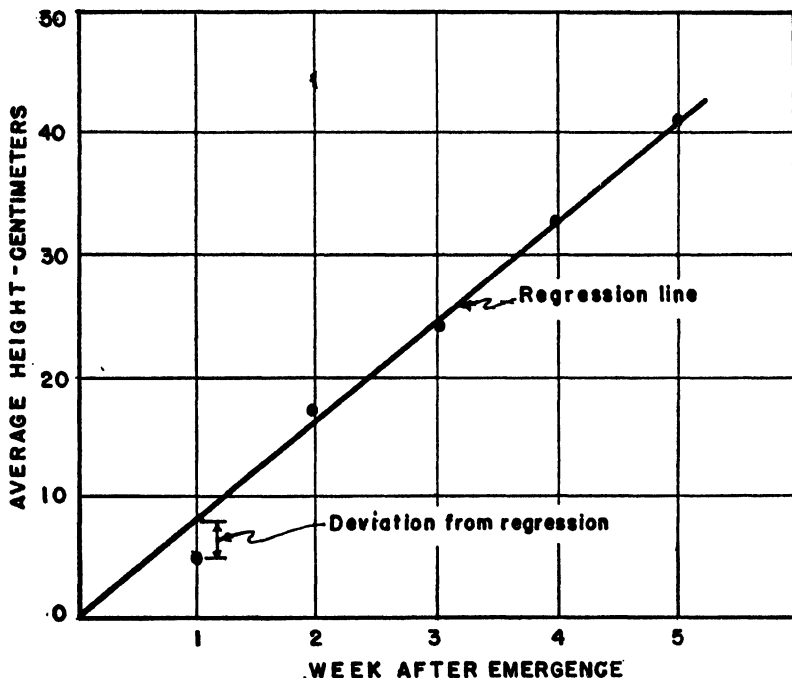


FIG. 6.1—Height of soybean plants in 5 successive weeks after emergence.

It is clear from both table and figure that growth was not uniform. During the several weeks the increases in average height were

5, 12, 7, 9, 8 centimeters.

Two questions present themselves: (i) Does this sample constitute evidence for uniform growth at this stage in the life of soybean plants or for a growth curve different from a straight line? (ii) If uniform growth is postulated, what is the best way to evaluate the average growth rate? The latter question will be attacked first because it is easier. For the time being we shall assume that this series of samples may have been drawn from a population in which there is a linear relation between age and height. That doesn't mean that every plant in the population is supposed to have a specified height at each age. The assumption is that the average

height increases uniformly with age. At any particular age the individual heights presumably vary about their mean in much the same way as did the weight gains of section 3.2, and the means of the weekly samples fall below or above the corresponding population values in the random manner characteristic of small sample averages. Question (ii) above may be rephrased thus: based on these samples, what is the best estimate of the population growth rate, assumed linear?

One might start by averaging the five weekly increases in height, the sum being 41 centimeters, the mean, 8.2 centimeters. But look again at that sum. It is clearly the average height attained by the last 4 plants. If this fifth group had averaged 45 centimeters instead of 41, it is easy to show that the mean of the 5 increases would have been $45/5 = 9$ centimeters instead of 8.2. So this average is completely determined by the 4 plants measured at the end of 5 weeks. An average dependent upon only a small part of the experimental data leaves much to be desired.

It might next be decided to calculate the growth rate to the end of each of the 5 weeks, then to average these. The rates would be $5/1 = 5$, $17/2 = 8.5$, $24/3 = 8$, $33/4 = 8.25$, and $41/5 = 8.2$ centimeters per week, and their average, 7.59. This would be an appropriate average if the standard deviation of the population of plants each week were proportional to age of the plants (5). While there is a tendency in the sample standard deviations to increase (observe the ranges), the trend is scarcely uniform enough to warrant setting up such an hypothesis. This averaging of growth rates gives most weight to the observations at the earlier ages.

Another way of arriving at an estimate of growth rate is known as the *method of averages*. The mean height of all the plants, $(5 + 17 + 24 + 33 + 41)/5 = 120/5 = 24$ centimeters, is divided by the mean age, $(1 + 2 + 3 + 4 + 5)/5 = 3$ weeks. The rate, $24/3 = 8$ centimeters per week, would be a proper estimate if the *variances* of the weekly samples were proportional to age. Again, the evidence is rather meager. This method lessens the weight ascribed to the heights near the beginning of growth.

A third method, one that seems more suitable for this sample, is based on the hypothesis that the weekly standard deviations in the population are all equal. This method will be presented in detail because of its greater utility, though it is clear that either of the other hypotheses might be acceptable so far as our data are concerned. The computation for this third method, somewhat more elaborate than those above, is described in the next section.

6.3—Calculation of regression passing through the origin. Homogeneous variance in the groups. The calculation set out in table 6.2 involves the products of the pairs of variates, X = number of weeks, Y = average height, as well as the square of X . (X is known as the independent variate or abscissa, while Y is the dependent or ordinate.) The formula for average growth rate, SXY/SX^2 , shows that a height, Y , is the more effective in determining growth rate the greater its corresponding

TABLE 6.2
CALCULATION OF REGRESSION THROUGH ORIGIN. HOMOGENEOUS VARIANCE
IN THE GROUPS

Week X	X^2	Average Height Y	XY	Estimates $\hat{Y} = 8.15X$	Deviations $Y - \hat{Y} = d_{y..}$
1	1	5	5	8.15	-3.15
2	4	17	34	16.30	0.70
3	9	24	72	24.45	-0.45
4	16	33	132	32.60	0.40
5	25	41	205	40.75	0.25
15	55	120	448		

Average growth rate = $SXY/SX^2 = 448/55 = 8.15$ cm./week.

age, X . In other words, the effect of each Y is augmented by its X , so that heights of later ages are more influential than those earlier. Technically speaking, under the present hypothesis the *weight*, which is inversely proportional to the weekly variance, is the same at all ages.

The rate, SXY/SX^2 , is known as the *regression coefficient*, ordinarily symbolized by b . Under the conditions already specified, it is an unbiased estimate of the population regression coefficient, β . One condition met in this example but not hitherto mentioned is that the sample values of X must be measured exactly. Errors in measuring X would bias the estimate, b . As heretofore, the term "error" does not refer to mistakes in measurement, though such would be included. Implied, rather, is the inability of the technician to make exact measurements. Practically, the age in weeks can be measured without error. But suppose the timepiece varied by as much as 24 hours forwards and backwards, then "error" would be introduced into the measurement of age, and the estimate of growth rate would be thereby biased, the tendency being to underestimate it.

Yet one more assumption is being made. Though X may be chosen at will, there must be no selection of Y ; it must be taken at random from all the Y 's corresponding to a given X . In our example, the time for measuring heights was arbitrary. It need not even have been spaced at equal intervals. But once the date of measurement was fixed, the height of the pre-selected plant must be recorded without prejudice. Another point of view is this: if the investigator wishes to select values for one of his variates, he thereby identifies it with X . The other variate becomes the random Y . If, in addition, values of Y were chosen, regression would cease to have meaning except perhaps as an evaluation of the whim of the chooser.

Since b is an estimate of the increase in height per week, the average population height at any age, X , is estimated as

$$\hat{Y} = bX = 8.15X \text{ centimeters}$$

This is known as the *regression equation* or the *regression of Y on X* . The symbol, $\hat{Y}$, may be read, *regression Y* or *estimated Y* or *predicted Y* . If any age, such as 3 weeks, is substituted for X , the corresponding value,

$$\hat{Y} = 8.15(3) = 24.45 \text{ centimeters,}$$

estimates the average population height for that age. In the table are estimates at each tabular date. Naturally, the sample averages deviate somewhat from the regression estimates, the differences, $Y - \hat{Y}$, being known as *deviations from regression* or *errors of estimate*. These differences will be symbolized by $d_{y.x}$.

If desired, estimates of height may be interpolated between whole weeks. Thus, on the 25th-day $X = 3.57$ weeks and

$$\hat{Y} = 8.15(3.57) = 29.1 \text{ centimeters}$$

One can even extrapolate on the assumption that the same linear growth rate will continue. At age 6 weeks, the height is then estimated as

$$\hat{Y} = 8.15(6) = 48.9 \text{ centimeters}$$

6.4—The regression line. This graph of the regression equation is the straight line drawn in figure 6.1. By hypothesis it passes through (0, 0), and is constructed by plotting any or all of the points (1, 8.15), (2, 16.30), etc. The growth rate, 8.15, is referred to as the *slope* of the line because this line rises 8.15 centimeters with each week. Deviations from regression are depicted clearly by vertical distances from the line to the observation points, positive deviations indicating points above the line and negative, below. These deviations measure the failure of the sample points to conform to the estimated trend. Notice the large negative deviation for the first week. One often sees this. Apparently it characterizes the transition from use of nutrients stored in the seed to those elaborated by the new plant.

The regression line is a kind of moving average, passing among the points that determine it in a manner specified by the particular hypotheses set up. For a given X , such as 3 weeks, the regression value of Y , 24.45 centimeters, is a more accurate estimate of the population height than is the sample value, 24 centimeters. One reason is that the regression estimate is based on 20 observations instead of the weekly 4. Another is that the height reached in any one week might be affected by unusual environmental conditions which tend to be equalized over the 5-week period.

Let us summarize the procedure outlined above. If you are fitting a regression through the origin, first examine the group variances to learn if they are homogeneous. For a test of this hypothesis, see section 10.13. If you decide that the population variances may be homogeneous, use the method of table 6.2. If you reject that hypothesis, first plot the standard deviations against X , then the variances. If the latter seem to yield a line approximately straight, use the method of averages; if the former,

use the first method given, averaging the weekly slopes, Y/X . Verify the facts that X is measured without appreciable error and that the corresponding Y 's are random.

EXAMPLE 6.1—Following are measurements on the heights of soybean plants in the field (17).

Age in weeks	1	2	3	4	5	6	7
Height in centimeters	5	13	16	23	33	38	40

The weekly variance seemed uniform throughout the period. Calculate the regression, $\hat{Y} = 6.029X$, by the method of table 6.2. Graph the data and plot the line. Compute the deviations from regression, $-1.03, 0.94, -2.09, -1.12, 2.86, 1.83$, and -2.20 .

EXAMPLE 6.2—In the text example, assume the standard deviation in the population to be proportional to age, so that $b = 7.59$ centimeters per week. Calculate the deviations from regression, $-2.59, 1.82, 1.23, 2.64$, and 3.05 . Note the large effect of the first week's height in making small the slope of the regression line.

EXAMPLE 6.3—Compute the deviations from regression in the text example if the method of averages is used. Ans. $-3, 1, 0, 1$ and 1 . Note the increasing emphasis on the later heights: while the first deviation is larger all the later ones are smaller.

EXAMPLE 6.4—Dungan (6) studied the effect on yield consequent upon the removal of various fractions of the leaves of maize plants in Illinois. In one experiment the leaves were stripped off when the grain was in the milk stage, the total leaf area being reduced by $16\frac{1}{2}\%$, $33\frac{1}{2}\%$, 50% , $66\frac{1}{2}\%$, $83\frac{1}{2}\%$ and 100% in six lots of plants. The corresponding reductions in the yield of shelled grain were 13, 17, 28, 22, 37, and 50%. Assuming a regression line through the origin and variance proportional to age, use the method of averages to compute the regression line, $\hat{Y} = 0.477X$. Do you suppose the values of X were measured without error?

6.5—The regression, $\hat{Y} = a + bX$. Homogeneous variance. X measured without error. Y randomly taken. Regressions passing through the origin are somewhat unusual. Ordinarily a more suitable equation is of the form, $\hat{Y} = a + bX$, where b represents the slope as before and a is the ordinate of the point where the line crosses the Y axis ($X = 0$). For comparison we shall fit this more adaptable line to the soybean data. The computations in table 6.3 lead to the evaluation of the two constants, a and b . With the exception of the last column, all the calculations follow a pattern which has become familiar in earlier chapters. In the last column are products of corresponding x - and y -deviations, such as $(-2)(-19) = 38$. The sum of these products divided by the number of degrees of freedom,

$$\frac{S_{xy}}{n-1} = \frac{88}{4} = 22,$$

is known as the covariance of X and Y .

The regression coefficient is given by

$$b = S_{xy}/S_x^2 = 88/10 = 8.8 \text{ cms. per week.}$$

TABLE 6.3
CALCULATION OF LINEAR REGRESSION. X MEASURED WITHOUT ERROR AND
VARIANCE HOMOGENEOUS
Mean Heights of Soybean Plants

Age (weeks)	Height (cms.)	Deviations From Mean		Squares of Deviations		Products of Deviations
		x	y	x^2	y^2	
X	Y					xy
1	5	-2	-19	4	361	38
2	17	-1	-7	1	49	7
3	24	0	0	0	0	0
4	33	1	9	1	81	9
5	41	2	17	4	289	34
Sum	15	120		10	780	88
Mean	3	24				

$$\hat{Y} = \bar{y} + \frac{S_{xy}}{S_x^2}(X - \bar{x}) = 24 + \frac{88}{10}(X - 3) = 24 + 8.8X - 26.4 = -2.4 + 8.8X \text{ cms.}$$

Compare this with the formula of section 6.3 where deviations from means were not used.

No attention is now being paid to the origin of coordinates, but there is another point, $(\bar{x}, \bar{y})$, marked by an arrow in figure 6.2, which assumes

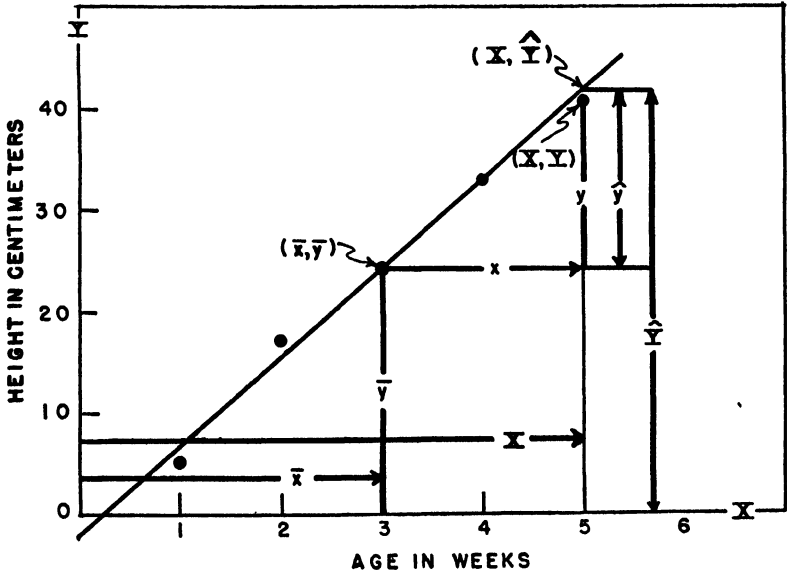


FIG. 6.2—Regression line fitted to soybean growth data. Line passes through origin of deviations, (3, 24), and cuts Y axis 2.4 centimeters below origin of coordinates.

the same role in this present regression method. It may be called the origin of deviations, since both x and y are measured from it in all four directions. The regression now passes through this origin. To the right of $(\bar{x}, \bar{y})$ the line rises 8.8 centimeters for each week in excess of $\bar{x} = 3$, so that its equation may be written,

$$\hat{y} = 8.8x \text{ centimeters}$$

As an example of the use of regression, we may estimate the population deviation from mean height at age, $x = 2$, (that is, $X = 3 + 2 = 5$ weeks) as

$$\hat{y} = 8.8(2) = 17.6 \text{ centimeters}$$

Compare this with the sample deviation, $Y - \bar{y} = 41 - 24 = 17$ centimeters. Note that $\hat{y} = \hat{Y} - \bar{y}$ (figure 6.2).

Instead of using deviations, we may return to the original measurements by substituting $X - \bar{x}$ for x and $\hat{Y} - \bar{y}$ for $\hat{y}$. Then

$$\hat{Y} - \bar{y} = 8.8(X - \bar{x})$$

or

$$\hat{Y} - 24 = 8.8(X - 3) \text{ centimeters}$$

for the soybean heights. According to the convenience of the user, this equation may be written as

$$\hat{Y} = 24 + 8.8(X - 3)$$

or

$$\hat{Y} = -2.4 + 8.8X \text{ centimeters.}$$

The advantage of this latter form is its brevity. A disadvantage is that the constant term, -2.4 , is a hodgepodge of two means and a regression coefficient with no clear biological meaning. If $X = 1$ week is now substituted in either equation, $\hat{Y}$ is found to be 6.4 centimeters, and the deviation from regression, $Y - \hat{Y} = 5 - 6.4 = -1.4$ centimeters. The other values of $\hat{Y}$ and $Y - \hat{Y}$ are recorded in table 6.4, and from them (or any two of them) the line in the figure is plotted.

TABLE 6.4

REGRESSION ESTIMATES OF SOYBEAN HEIGHTS CONTRASTED WITH OBSERVED HEIGHTS

Age X	Average Height Y	Regression Estimate $\hat{Y}$	Deviation From Regression $Y - \hat{Y} = d_{y \cdot x}$	Square of Deviation $(Y - \hat{Y})^2 = d_{y \cdot x}^2$
1	5	6.4	-1.4	1.96
2	17	15.2	1.8	3.24
3	24	24.0	0.0	0.00
4	33	32.8	0.2	0.04
5	41	41.6	-0.6	0.36
Sum	120	120.0	0.0	5.60

There are now apparent two notable features of this type of regression. First, the sum of $\hat{Y}$ being the same as the sum of Y , it follows that the sum of the deviations from regression is zero. This reminds one of the fact that the sum of deviations from the arithmetic mean is zero. Thus, the new average, $\hat{Y}$, has this distinctive characteristic of $\bar{y}$: the regression line passes among the sample points in such manner that the sum of the deviations therefrom is zero; that is, the sum of the positive deviations from points above the line is equal to the sum of the negative from points below.

The second feature is that $S(Y - \hat{Y})^2 = 5.60$ is far less than the sum of the squares of deviations from mean, $Sy^2 = 780$. In fact, under our present hypotheses (homogeneous variance, X measured without error) 5.60 is less than the sum of squares of deviations from any other straight line. We are again struck by the analogy between $\hat{Y}$ and $\bar{y}$ since the latter average also is situated among the sample points in such fashion that Sy^2 is a minimum. Reference to figure 6.2 will make this clearer. The sum of the squares of deviations (780) from the line, $\bar{y} = 24$, is less than that from any other *horizontal* line; whereas the sum of squares, 5.60, is less than that from any other line (compare example 6.9).

EXAMPLE 6.5—On alternate days, random individuals were measured, as follows:

Individual number	1	2	3	4	5	6	7
Day, X	2	4	6	8	10	12	14
Weight, Y	4	2	5	9	3	11	8

Compute the regression, $\hat{Y} = 2 + 0.5X$, and the deviations from regression, 1, -2, 0, 3, -4, 3, and -1. Graph the data and verify your results by measuring the deviations.

EXAMPLE 6.6—The following weights of body and comb of 15-day-old White Leghorn male chicks are adapted from Snedecor and Breneman (13):

Chick Number	1	2	3	4	5	6	7	8	9	10
Body weight (grams), X	83	72	69	90	90	95	95	91	75	70
Comb weight (milligrams), Y	56	42	18	84	56	107	90	68	31	48

Calculate the regression, $\hat{Y} = 60 + 2.302(X - 83)$.

EXAMPLE 6.7—In the preceding example, calculate the sum of the squares of the deviations from regression. Ans. 1,554.8. Which comb weight deviates most from regression?

EXAMPLE 6.8—Construct the graph of the chick data, plotting body weight along the horizontal axis. Insert the regression line. Do the deviations from regression verify the computed values?

EXAMPLE 6.9—Compare $S(Y - \hat{Y})^2 = 5.60$ (table 6.4) for the regression $\hat{Y} = a + bX$ with the similar sum of squares, 10.84, for the regression through the origin. To get this latter figure, add the squares of the deviations in the last column of table 6.2.

TABLE 6.5
REGRESSION OF PERCENTAGE OF WORMY FRUIT ON SIZE OF APPLE CROP

Tree Number	Size of Crop on Tree (hundreds of fruits)	Percentage of Fruits Wormy		Deviations From Mean		Squares of Deviations		Products of Deviations	Regression Percentage	Deviation From Regression	Square
		X	Y	x	y	x^2	y^2				
1	8		59	-11	14	121	196	-154	56.14	2.86	8.18
2	6		58	-13	13	169	169	-169	58.17	-0.17	0.03
3	11		56	-8	11	64	121	-88	53.10	2.90	8.41
4	22		53	3	8	9	64	24	41.96	11.04	121.88
5	14		50	-5	5	25	25	-25	50.06	-0.06	0.00
6	17		45	-2	0	4	0	0	47.03	-2.03	4.12
7	18		43	-1	2	1	4	2	46.01	-3.01	9.06
8	24		42	5	-3	25	9	-15	39.94	2.06	4.24
9	19		39	0	-6	0	36	0	45.00	-6.00	36.00
10	23		38	4	-7	16	49	-28	40.95	-2.95	8.70
11	26		30	7	-15	49	225	-105	37.91	-7.91	62.57
12	40		27	21	-18	441	324	-378	23.73	3.27	10.69
Sum	228		540	0	0	924	1,222	-936	540.00	-0.00	273.88
Mean	19		45						45.00		

$$\hat{Y} = \bar{y} + \frac{S_{xy}}{S_x^2}(X - \bar{x}) = 45 - \frac{936}{924}(X - 19) = 64.247 - 1.013X \text{ hundreds of fruits}$$

Release from the restriction that the regression contain the origin has resulted in a better fit of the line to the points, in the sense that the less the sum of squares the better the fit.

Extension of this comparison to the deviations from the first two lines considered in section 10.2 involves the introduction of *weights*. For the first method, the weights are inversely proportional to the squares of X so that the weighted sum of squares of the deviations in example 6.2 is

$$\frac{(-2.59)^2(1/1^2) + (1.82)^2(1/2^2) + \dots + (3.05)^2(1/5^2)}{(1/1^2) + (1/2^2) + \dots + (1/5^2)} = 5.82$$

This makes it plain that the deviation, -2.59 , gets the most weight while 3.05 gets the least.

In the method of averages the weights are inversely proportional to X , the weighted sum of squares of the deviations (example 6.3) being

$$\frac{(-3)^2(1/1) + (1)^2(1/2) + \dots + (1)^2(1/5)}{(1/1) + (1/2) + \dots + (1/5)} = 4.36$$

This indicates that among the various sets of hypotheses considered, the pair specifying (i) regression containing the origin and (ii) variance proportional to X results in the best *sample* fit.

6.6—Regression of injured fruit on crop size. It seems wise to interrupt the discussion at this point to introduce a new illustration, one in which the regression obviously does not contain the origin of coordinates.

It is rather generally thought that the intensity of the injury by codling moth larvae is greater on apple trees bearing a small crop. Apparently the density of the flying moths tends towards uniformity, so that the chance of attack for any particular fruit is augmented if there are few fruits in the tree. The data in table 6.5 are adapted from the results of an experiment (9) containing evidence about this phenomenon. The 12 trees were all given a calyx spray of lead arsenate followed by five cover sprays made up of 3 pounds of manganese arsenate and 1 quart of fish oil per 100 gallons. There is a decided tendency, emphasized in figure 6.3, for the percentage of wormy fruits to decrease as the number of apples in the tree increases. In this particular group of trees, the relation of the two variates is even closer than usual.

The new feature in the calculations is the majority of negative products, xy , caused by the tendency of small values of Y to be associated with large values of X . The regression coefficient shows that the estimated percentage of wormy apples decreases, as indicated by the minus sign, 1.013 with each increase of 100 fruits in the crop. The regression line, and of course the percentage, falls away from the point, $A(\bar{x}, \bar{y})$, by 1.013 for each unit of crop above 19 hundreds.

The regression line brings into prominence the deviations from this moving average, deviations which measure the failure of crop size to account for variation in the intensity of infestation. Trees number 4, 9, and 11 had notably discrepant percentages of injured fruits, while numbers 2 and 5 performed as expected. Presumably these are random deviations from the average (regression) values, but close observation of the trees during the flight of the moths might reveal some characteristics of this

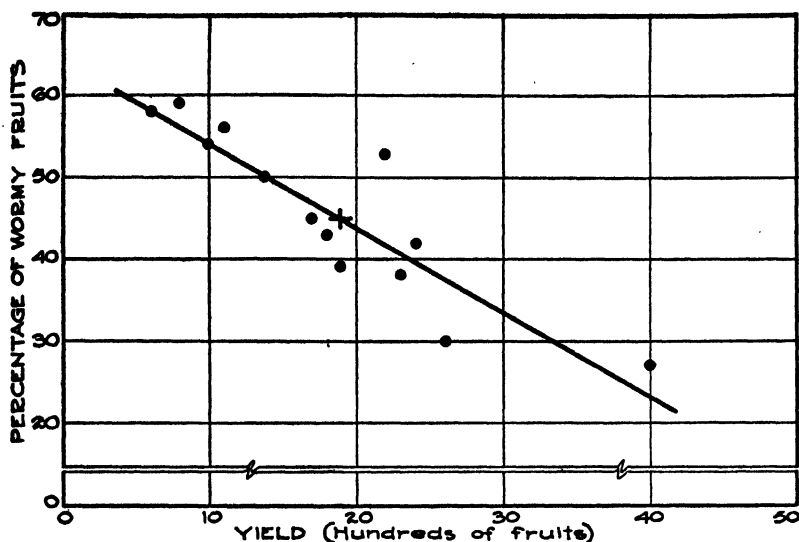


FIG. 6.3—Regression of percentage of wormy fruits on size of crop in apple trees. The cross indicates the origin for deviations, $A(\bar{x}, \bar{y})$.

phenomenon. Tree 4 might have been on the side from which the flight originated or perhaps its shape or situation caused poor applications of the spray. Trees 9 and 11 might have had some peculiarities of conformation or foliage that protected them. Careful study of trees 2 and 5 might throw light on the kind of tree or location that receives normal infestation. This kind of *case study* usually does not affect the handling of the sample statistics, but it may add greatly to the investigator's knowledge of his experimental material and may afford clues to the improvement of future experiments.

There are two extremes among attitudes toward experimental data, both of which should be avoided. Some attend only to minute details of sample variation, neglecting the summarization of the data and the consequent inferences about the population. Others are impatient of the data themselves, rushing carelessly toward averages and other generalizations. Either course fails to yield full information from the experiment. The competent investigator takes time to examine each datum together with the individual measured. He attempts to distinguish normal variation from aberrant observations. He then judges his summaries and his population inferences and draws his conclusions against this vivid background of sample facts.

6.7—Comparison of individuals and of group means. Adjusted values of Y . You have doubtless been impressed by the fact that when there is regression the statement, "One tree had only 27% wormy fruits

in an orchard whose mean was 45%," may be inadequate. You immediately ask, "What was the size of the tree crop?" If you learn that it was 40 bushels you become aware that 27% was a greater rate of infestation than one would expect in a tree with so large a yield. It is not the orchard average with which the tree percentage should be compared, but it is the expected percentage for the individual crop size. The regression, in fact, replaces the mean as a standard of comparison. This new standard is not the same for all trees but varies according to the number of apples borne.

In the case studies mentioned in the foregoing section two individuals are compared, not directly by their percentages but by contrasting their deviations from regression. The rates of infestation of two trees are immediately comparable only if their yields are the same. Otherwise, it is their errors of estimate that must be compared. From table 6.5 it is seen that, while trees 1 and 4 differ by $59 - 53 = 6\%$ in actual percentages of wormy fruits, yet when they are compared on the basis of crop size the difference is $2.86 - 11.04 = -8.18\%$, meaning that if these trees had borne the same number of apples the percentage of injured fruits on 1 is estimated to have been less than that on 4 by 8.18%. On the basis of yield one would expect these trees to differ by $56.14 - 41.96 = 14.18\%$, whereas they actually differ by only 6% ($6 - 14.18 = -8.18\%$).

Instead of comparing deviations from regression it is often more convenient first to *adjust* all values of Y to a common X (usually $\bar{x}$), then compare the adjusted percentages. This adjusted Y is $\bar{y}$ plus the deviation from regression (figure 6.4). To table 6.5 one could easily append a

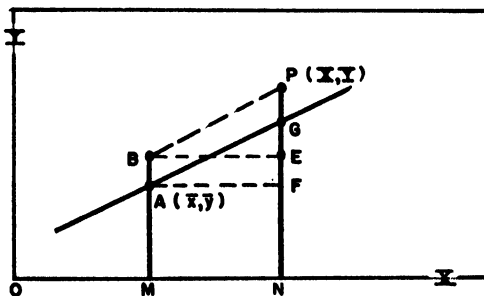


FIG. 6.4—Regression of Y on X . $NP = Y$ and $MB = (\text{adjusted } Y) = \bar{y} + AB = \bar{y} + GP = \bar{y} + (\text{deviation from regression})$. Also, $MB = NE = Y - EP = Y - FG = Y - b(AF) = Y - bx$.

column containing the 12 adjusted percentages, 47.86, 44.83, 47.90, 56.04, etc. These represent the rates of infestation that might have been expected if the trees had all borne the mean number of apples, 1,900. Differences among these adjusted values of Y are identical with corresponding differences among the deviations from regression. Thus, you

see that the adjusted percentage for tree 4 is greater than that for tree 1 by $56.04 - 47.86 = 8.18\%$.

It isn't necessary to calculate deviations from regression in order to arrive at adjusted percentages. The latter may be got directly by use of the formula,

$$\text{Adjusted } Y = Y - bx,$$

where b is the regression coefficient and x is deviation from mean, as usual. For example, take tree 1 in the table. Since $Y = 59$, $b = -1.013$ and $x = -11$, the adjusted percentage is

$$59 - (-1.013)(-11) = 47.86\%,$$

which checks with the result given above.

Not only individuals, but also the means of two groups of percentages may be compared on the basis of their deviations from regression (or their adjusted values). The pertinent hypothesis is that the two groups have the same population regression, β , and that b is its estimate. As an example, suppose the first four apple trees to be of variety A and the last four, B . Then the difference between their mean deviations from regression is

$$(2.86 - 0.17 + 2.90 + 11.04)/4 - (-6.00 - 2.95 - 7.91 + 3.27)/4 = 7.56\%,$$

the sample trees of variety B being less susceptible to attack.

EXAMPLE 6.10—Calculate adjusted Y for each of the eight trees of varieties A and B in the paragraph just above, then show that the difference between the mean adjusted percentages of A and B is 7.56% .

EXAMPLE 6.11—You can get the same answer as in the foregoing example by doing this: calculate the mean Y and mean x in each group, adjust each mean Y , then subtract. Can you still further shorten the process? These short cuts are due to the linearity of the regression.

EXAMPLE 6.12—Another group of 12 trees, investigated by Hansberry and Richardson, was sprayed with lead arsenate throughout the season. In addition, the fourth and fifth cover sprays contained 1% mineral oil emulsion and nicotine sulfate at the rate of 1 pint per 100 gallons. The results are shown below. These facts may be verified: $SX = 240$, $SY = 384$, $Sx^2 = 808$, $Sy^2 = 1428$, $Sxy = -582$, regression coefficient $= -0.7203$, $\bar{Y} = 46.41 - 0.7203\bar{X}$, $Y - \bar{Y}$ for the first tree $= 16.40\%$.

Size of Crop, X Apples	15, 15, 12, 26, 18, 12, 8, 38, 26, 19, 29, 22
Percentage Wormy, Y	52, 46, 38, 37, 37, 37, 34, 25, 22, 22, 20, 14

EXAMPLE 6.13—From the data on 98 lots of 6 pigs each the following results were computed:

Variate	Symbol	Mean	Variance	Covariance
Pounds of feed consumed per lot	X	4,669.0	899,557	
Pounds gain in weight per lot	Y	1,180.7	39,482	178,959

Verify the regression equation, $\hat{Y} = 0.19894X + 251.8$. Do you see why the regression coefficient is carried to five significant figures?

EXAMPLE 6.14—Among the lots of pigs in the preceding example, lot A, consuming 6,863 pounds of feed, gained 1,612 pounds; while lot B, with 3,688 pounds of feed, gained 996 pounds. Are the gains of these two lots in accord with expected gains? Represent graphically the regression equation, and plot points for A and B.

EXAMPLE 6.15—In lot C of the pigs above, $X = 6,426$ and $Y = 1,380$ pounds. In lot D, $X = 4,846$ and $Y = 1,362$ pounds. Compute the errors of estimate for these lots. Do you agree that the net difference between lot gains, after feed consumption is compensated, is 296.3 pounds? Plot points corresponding to C and D, and observe the graphical interpretation of this difference. Observe that Y is almost the same in the two lots.

EXAMPLE 6.16—The 98 lots of example 6.13 were grouped in various experiments conducted at various times and under various environmental conditions. In one of these experiments containing 17 lots, the average amount of feed per lot was 4,007 pounds, with average lot gain 1,087 pounds. In a second with 13 lots, the figures were 6,565 pounds of feed and 1,546 pounds gain. Verify the fact that the net difference in gain, after proper adjustment for feed consumption, averaged 50 pounds per lot.

EXAMPLE 6.17—Could you have got the estimated difference in the last example by using only the regression coefficient, 0.19894 pound gain per pound feed consumed, to reach the correct result?

6.8—Standard deviation from regression (or standard error of estimate). The deviations from regression are squared and then added in each of the tables 6.4 and 6.5. Attention was called to the fact that, under each hypothesis set up, this sum of squares is less than any similar sum that can be got from other straight lines passing among the datum points. In other words, the methods of calculation that we have adopted follow from the specification that the sum of the squares of deviations from linear regression shall be the minimum. One part of the theory provides that these deviations shall be randomly assorted, falling in the table without any predetermined pattern. This is another way of saying that the standard deviation in the population is the same for all values of X and is one distinction between the present method of fitting and others such as the first two described in section 6.2.

For simplicity, the deviation from regression, $Y - \hat{Y}$, has been denoted by $d_{y \cdot x}$, indicating the deviation of Y from the regression of Y on X . Then the mean square of the deviations from regression,

$$\frac{Sd_{y \cdot x}^2}{n - 2} = \frac{273.88}{12 - 2} = 27.388 \text{ (applied to the apple data),}$$

that is, the variance of the deviations, is an unbiased estimate of the population variance from regression. Its square root, $\sqrt{27.388} = 5.23\%$, the standard deviation from regression, is an average of the deviations from regression in the next-to-last column of table 6.5. This is an estimate of that population standard deviation of Y which is the same at all values of X . It will be observed that the sum of the deviations from regression is zero so that there is no correction for mean to be subtracted from 273.88. The new expression for degrees of freedom, $n - 2$, results from the fact that *two* averages involving Y have been taken from the sample data, first the mean, $\bar{y}$, and now the regression coefficient, b . For each of them one degree of freedom must be deducted from n . If division were by any other number than $n - 2$, the estimate of variance would be biased.

The standard deviation from regression is symbolized by $s_{y \cdot x}$, denoting the variation in Y after its regression on X has been discounted. In the apple data, $s_{y \cdot x} = 5.23\%$ measures the amount of variation (in percentage injured apples) not associated with, explained by, or dependent upon changing amounts of tree yield. Interpreting it graphically, this is an average of the vertical distances of the dots (figure 6.3) from the regression line.

It isn't necessary to follow the calculations in the last three columns of table 6.5 in order to get $Sd_{y \cdot x}^2$. It may be shown (example 6.22) that

$$Sd_{y \cdot x}^2 = S_y^2 - (S_{xy})^2/S_x^2 = 1,222 - (-936)^2/924 = 273.84,$$

which is a more nearly exact figure than the 273.88 in the table because the latter is worse affected by rounding errors.

EXAMPLE 6.18—Applying the formula immediately above to the results in table 6.3, calculate $Sd_{y \cdot x}^2$ and compare the answer with table 6.4.

EXAMPLE 6.19—From the result of the preceding example, calculate $s_{y \cdot x} = 1.366$ cms. This is an average of what deviations?

EXAMPLE 6.20—In the chick-hormone data of example 6.7 you had $Sx^2 = 1,000$, $Sy^2 = 6,854$, $S_{xy} = 2,302$. Calculate $s_{y \cdot x} = 13.9$ milligrams.

EXAMPLE 6.21—From the soybean heights in table 6.1 there are available two independent estimates of the population variance, assumed constant at all the sample ages. The first is got by averaging 5 variances, one calculated from the heights of the four plants measured each week. The variance for the first week is $s^2 = 3.33$. Compute the other four, together with the mean of all five, 8.67. The second estimate is made from the variance-from-regression of the 5 weekly means (example 6.19). This is $4s_{y \cdot x}^2 = 4(5.6/3) = 7.47$. Multiplication is by 4 because $s_{y \cdot x}^2$ is the variance of means, only one-fourth of the variance of plants. In chapters 10 and 11 will be found more adequate methods for treating these soybean data.

EXAMPLE 6.22—Those interested in algebra will profit by deriving the formula, $Sd_{y \cdot x}^2 = S_y^2 - (S_{xy})^2/S_x^2$. The easy way is to work with deviations from mean,

$$d_{y \cdot x} = y - \hat{y} = y - bx$$

Square both sides of the equation, then sum, remembering that $b = S_{xy}/S_x^2$.

6.9—Fiducial statements and tests of null hypotheses. Having in hand an unbiased estimate of $\sigma_{y \cdot x}^2$, the variance from regression in the

population, we are able to proceed with probability statements about the various averages of regression.

First, there is the regression coefficient, b , an estimate of the population coefficient, β . In random sampling from this population, b is distributed with a variance estimated by

$$s_b^2 = s_{y \cdot x}^2 / Sx^2$$

Thus, in the apple data,

$$s_b^2 = 27.388/924 = 0.0296, s_b = 0.172\%$$

Moreover, the quantity $(b - \beta)/s_b$ follows the t -distribution with $n - 2$ degrees of freedom. Hence, fiducial limits may be set,

$$l_1 = b - t_{.05} s_b \quad \text{and} \quad l_2 = b + t_{.05} s_b, \quad (t_{.05} = 2.228, d.f. = 10)$$

which, for the apples, are $-1.013 - 0.383 = -1.396$ and $-1.013 + 0.383 = -0.630$ percent per 100 fruits. If we say that the population regression is within these limits we shall be right unless our sample is one of the divergent kind that occurs about once in 20 trials. But perhaps it is our interest to test the hypothesis that $\beta = 0$ (or some other value that may seem reasonable). Then

$$t = \frac{b - \beta}{s_b} = \frac{1.013}{0.172} = 5.89 \text{ for the apples.}$$

Certainly this would be a remarkably aberrant result if the sampling were from the hypothetical population in which $\beta = 0$. One would probably reject that hypothesis, then conclude that in the population there is a regression of percentage wormy apples on crop size, the value likely being between -0.630 and -1.396 percent per 100 fruits.

Next consider the mean. In unrestricted sampling its variance is, as usual, s^2/n . But in random sampling of both X and Y , as well as in sampling from a population confined to the *same values of* X as occurred in the sample under consideration (these values of X may be chosen in advance of the sampling if desired, as they were in the soybean experiment) the opportunity for variation is less, the variance being $s_{y \cdot x}^2/n = 27.388/12 = 2.282$ for the apples. Thus, $\sqrt{2.282} = 1.51\%$ is the standard error of the mean, 45%, under the circumstances stated. Fiducial limits may be set: $l_1 = 45 - (2.228)(1.51) = 41.6\%$ and $l_2 = 45 + (2.228)(1.51) = 48.4\%$. Furthermore, any relevant null hypothesis may be tested. Clearly $m = 0$ is fantastic, but it might have been known that over a large fruit growing region in the same year the damage to apples was 50%. To get evidence about the comparative rate of injury in this orchard, one might set up the hypothesis that this sample of apples was drawn from a population in which $m = 50$. Then

$$t = (45 - 50)/1.51 = 3.31 \text{ (in absolute value)}$$

Since this value of t is outside the 1% level, 3.169, the mean of 45% would

be unusually divergent if the sampling were from the hypothetical $m = 50\%$ with trees having crop sizes from 800 to 4,000 apples. It might be concluded that (i) the spraying done was more effective than the usual practice, or (ii) that this orchard was, for some unknown reason, less heavily attacked than others, or (iii) that it had more than the average number of apples per tree.

A third set of statements may be made about any $\hat{Y}$ which may be of interest. For the average crop size, 1,900 apples, the mean percent injury is estimated as $\hat{Y} = \bar{y}$, the case just discussed. For any other X the estimate, $\hat{Y}$, is subject to an additional source of variation. Not only may the mean, $\bar{y}$, be in error but also the regression, b . Since $\hat{Y} = \bar{y} + bx$, its variance is the sum of the two independent variances of $\bar{y}$ and of bx . This is

$$s_{\hat{Y}}^2 = \frac{s_{y \cdot x}^2}{n} + \frac{s_{y \cdot x}^2 x^2}{Sx^2} = s_{y \cdot x}^2 \left(\frac{1}{n} + \frac{x^2}{Sx^2} \right)$$

Hence,

$$s_{\hat{Y}} = s_{y \cdot x} \sqrt{1/n + x^2/Sx^2}$$

It is clear that this standard error increases with x , that is, as X recedes from $\bar{x}$. The reason for this is rather obvious in a graphical representation of regression. Any error in the slope of the line has a more pronounced effect on $\hat{Y}$ the further removed it is from the center of the distribution. In the apple data,

$$s_{\hat{Y}} = \sqrt{27.388(1/12 + x^2/924)} = \sqrt{2.282 + 0.02964x^2}$$

Thus, for trees bearing crops like number 12, $x = 21$ and $s_{\hat{Y}} = 3.92\%$, notably greater than $s_y = 1.51\%$. In ordinary practice the second term is relatively small within the sample range, but in forecasting beyond this range the term with x^2 may become, as above, predominant. If we now say that for trees bearing 4,000 fruits the population mean percentage of injured is between $23.73 - (2.228)(3.92) = 15.00\%$ and $23.73 + (2.228)(3.92) = 32.46\%$, the probability of being right is 95%.

One often wishes to make a fiducial statement about injury in a specific tree, Y , rather than about the corresponding population mean, $\bar{Y}$. For that kind of statement, the tree variance, $s_{y \cdot x}^2$, must be added, bringing the total to

$$s_{y \cdot x}^2(1 + 1/n + x^2/Sx^2)$$

This implies that while the population mean percentage for trees with 4,000 apples is probably between 15.00% and 32.46%, yet a randomly selected tree with that yield may be found to have anywhere between 9.16% and 38.30% injured fruits. You see, predicting the performance of individuals, especially of those in the outer fringe, is a hazardous venture unless $s_{y \cdot x}$ is unusually small.

For purposes of illustration we have made all the probability state-

ments above as though they were independent of each other. Statements about $\bar{y}$ and b are independent, but it is clear that probabilities related to $\hat{Y}$ and Y are determined in part by the same accidents of sampling that produced $\bar{y}$ and b . It is proper to consider the implications of all these statements. Only remember that if it should be found, for example, that the population mean lies beyond the sample based fiducial limits, then the probability would be much greater than .05 that the population value of regression Y is outside the fiducial limits based on $\hat{Y}$.

EXAMPLE 6.23—In the regression of comb weight on body weight, example 6.7, $n = 10$, $\bar{x} = 83$ grams, $\bar{y} = 60$ mg. $Sx^2 = 1,000$, $Sy^2 = 6,854$ and $Sxy = 2,302$. Set 5% fiducial limits on $\bar{y}$, assuming the same set of body weights. Ans. 49.8 — 70.2 mg.

EXAMPLE 6.24—In the chick data, $b = 2.302$. Test the hypothesis that $\beta = 0$. Ans. $t = 5.22$, $P < .01$.

EXAMPLE 6.25—Since evidently there is a population regression of comb weight on body weight, set 95% fiducial limits to the regression coefficient. Ans. 1.28 — 3.32 mg. per gram.

EXAMPLE 6.26—Predict the population average comb weight of 100-gram chicks. Ans. 99.1 mg. with 95% limits, 79.0 — 119.2 mg.

EXAMPLE 6.27—Set 95% fiducial limits to the forecast of the comb weight of a randomly chosen 100-gram chick. Ans. 61.3 — 136.9 mg.

6.10—Short-cut methods of computation in regression. So far in this chapter you have not been greatly burdened with computation. In practice, regression calculations are rather tedious so that either machines are used or some form of coding is practiced. Short cuts in regression are mainly those already explained in chapter 5, only a single new feature being introduced. The data in table 6.6 will serve as illustration.

TABLE 6.6
INITIAL WEIGHTS AND GAINS IN WEIGHT (GRAMS) OF 15 FEMALE RATS ON HIGH
PROTEIN DIET, 28TH TO 84TH DAYS OF AGE

	Rat Number														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Initial weight, X	50	64	76	64	74	60	69	68	56	48	57	59	46	45	65
Gain, Y	128	159	158	119	133	112	96	126	132	118	107	106	82	103	104

If a machine is available the methods of section 5.2 are easily extended to the calculation of regression. In most machines the sum of either variate is run up along with the sum of its squares; that is, both SX and SX^2 appear in the recording dials at the end of a single series of operations. As for the sum of products,

$$(50)(128) + (64)(159) + \dots + (65)(104) = 108,530,$$

it is also read from the dials after the sequence of multiplications, none of the individual products being recorded. Your book of instructions, or the nearest agent, will furnish directions. In some machines the two sums, the two sums of squares, and the sum of products may all be accumulated at one run. If there are available machines for tabulating data punched in cards, they will be found convenient for handling large samples (100 or more sets of observations) (1). The results from any of these machines are entered in a table like 6.7. One advantage of a machine is the avoid-

TABLE 6.7
COMPUTATIONS IN REGRESSION IF MACHINE IS USED. RAT DATA FROM TABLE 6.6

$SX = 901$	$SY = 1783$	$n = 15$
$\bar{x} = 60.07$ grams	$\bar{y} = 118.87$ grams	
$SX^2 = 55,465$	$SY^2 = 218,297$	$SXY = 108,530$
$(SX)^2/n = 54,120.07$	$(SY)^2/n = 211,939.27$	$(SX)(SY)/n = 107,098.87$
$Sx^2 = 1,344.93$	$Sy^2 = 6,357.73$	$Sxy = 1,431.13$

ance of much copying of numbers. This enhances both the accuracy and the rapidity of calculation. The only entries that need be made are those in table 6.7. Note the one new feature, the correction term for the sum of products. This contains the product of SX by SY instead of the square of either, rather a reasonable method of correcting products. Don't overlook the fact that the correction term may be larger than SXY , making Sxy negative. This would indicate a downward sloping regression line.

The six quantities,

$$n, \bar{x}, \bar{y}, Sx^2, Sy^2, Sxy,$$

constitute the entire information necessary for completing the computations of regression. We have, for example,

$$b = Sxy/Sx^2 = 1,431.13/1,344.93 = 1.0641$$

Gain in weight among such female rats is, therefore, partly predictable, each gram increase in initial weight corresponding to 1.0641 grams increase in gain. However,

$$Sd_{y \cdot x}^2 = Sy^2 - (Sxy)^2/Sx^2 = 6,357.73 - (1,431.13)^2/1,344.93 = 4,834.88,$$

$$s_{y \cdot x}^2 = Sd_{y \cdot x}^2/(n - 2) = 4,834.88/13 = 371.91,$$

$$s_b^2 = s_{y \cdot x}^2/Sx^2 = 371.91/1,344.93 = 0.2765$$

$$s_b = \sqrt{0.2765} = 0.526$$

$$t = b/s_b = 1.0641/0.526 = 2.02$$

The large value of s_b relative to b will be obvious if you graph the data in the manner of table 6.2. Since t does not quite reach the 5% level one must not be too confident about population facts. Nevertheless, the sample estimate cannot be ignored. There is a good deal of evidence to sup-

port some relation between initial weight and gain. We have here some information about experimental design and shall return to these results in the succeeding section.

The student is now prepared to read chapters 10 and 11, analysis of variance, or he may dip into chapters 7 and 9.

If a calculating machine is not available, one may profit by following directions for coding given in chapter 5. Each variate should be reduced to a series of integers centering on zero and having a range between 20 and 50 units. This not only makes the calculation relatively easy but also produces results sufficiently accurate for ordinary purposes. The process is illustrated in table 6.8. These data were taken from a study of the forces

TABLE 6.8
DRAFT AND SPEED OF PLOWS DRAWN BY TRACTORS
Illustration of coding in computation of regression

Speed Miles per Hour V	Draft Pounds D	Code Numbers $(10V - 30)/2$ (rounded) X	$(D/10) - 50$ (rounded) Y	Squares of Code Numbers X^2	Y^2	Products of Code Numbers XY
0.9	425	-10	-8	100	64	80
1.3	420	-8	-8	64	64	64
2.0	480	-5	-2	25	4	10
2.7	495	-2	0	4	0	0
3.4	540	2	4	4	16	8
3.4	530	2	3	4	9	6
4.1	590	6	9	36	81	54
5.2	610	11	11	121	121	121
5.5	690	12	19	144	361	228
6.0	680	15	18	225	324	270
34.5	5460	48	64	727	1044	841
		-25	-18			
		23	46			

$$SX = 23$$

$$\bar{x} = 2.3 \text{ code units}$$

$$\bar{v} = \frac{(2)(2.3) + 30}{10}$$

$$= 3.46 \text{ mi./hr.}$$

$$SX^2 = 727$$

$$(SX)^2/n = 52.9$$

$$Sx^2 = 674.1$$

$$Sv^2 = (674.1)(2/10)^2$$

$$= 26.964$$

$$SY = 46$$

$$\bar{y} = 4.6 \text{ code units}$$

$$\bar{d} = (4.6 + 50)(10)$$

$$= 546 \text{ pounds}$$

$$SY^2 = 1,044$$

$$(SY)^2/n = 211.6$$

$$Sy^2 = 832.4$$

$$Sd^2 = (832.4)(10)^2$$

$$= 83,240$$

$$SXY = 841$$

$$(SX)(SY)/n = 105.8$$

$$Sxy = 735.2$$

$$Svd = (735.2)(2/10)(10) = 1,470.4$$

necessary to draw plows at the speeds commonly attained by tractors (2). The character of the soil and depth of the furrow were experimentally controlled as closely as possible. Within the limits of the trial the draft was roughly proportional to the speed, as a graphical representation of the data will indicate.

In order to avoid large numbers and decimals the coding was carried to an extreme. Doubtless some will prefer the decimals and large numbers: an argument against coding is the opportunity for mistakes in the procedure itself. However, familiarity makes the process less forbidding, and there are data whose coding is certainly convenient.

The decoding offers some novel features. The sums of squares must be decoded by use of the squares of the coding ratios and the sum of products by means of their product. If this doesn't seem reasonable, try extracting the square roots of the sums of squares, decoding in the usual manner, then squaring the results.

The regression coefficient, $b = Svd/Sv^2 = 1,470.4/26.964 = 54.53$, shows that the drawbar pull increased an average of 54.53 pounds for each mile per hour increase in speed. The formula for estimating draft at various speeds is

$$\hat{D} = 546 + 54.53 (V - 3.46) = 54.53V + 357.3 \text{ pounds}$$

Since the use of this regression formula may lead to the prediction of drawbar pulls in individual trials, it is pertinent to calculate the variance, s_D^2 , of a randomly selected D :

$$\begin{aligned} S(D - \hat{D})^2 &= Sd^2 - (Svd)^2/Sv^2 = 83,240 - (1,470.4)^2/26.964 = 3,056 \\ s_{d \cdot v}^2 &= S(D - \hat{D})^2/(n - 2) = 3,056/8 = 382.0 \\ s_D^2 &= s_{d \cdot v}^2(1 + 1/n + v^2/Sv^2) = 382.0(1 + 0.1 + v^2/26.964) \\ &= 420.2 + 14.167v^2 \end{aligned}$$

As an example, let us estimate the drawbar pull of a tractor making 5 miles per hour.

$$\hat{D} = 54.53(5) + 357.3 = 630 \text{ pounds}$$

$$v = 5 - 3.46 = 1.54 \text{ miles per hour}$$

$$s_D^2 = 420.2 + 14.167(1.54)^2 = 453.8$$

$$\text{Standard error} = 21.3 \text{ pounds, } t_{.05} = 2.306$$

$$\text{Fiducial limits: } l_1 = 630 - (21.3)(2.306) = 581, l_2 = 679 \text{ pounds}$$

EXAMPLE 6.28—Speed records attained in the Indianapolis Memorial Day automobile races, 1911–1939 (miles per hour):

Year	1911	1912	1913	1914	1915	1916	1917	1918	1919	1920
Coded year, X	1	2	3	4	5	6	7	8	9	10
Speed, Y	74.7	78.7	75.9	82.5	89.8	83.3	(no races)		88.1	88.5

Year	1921	1922	1923	1924	1925	1926	1927	1928	1929	1930
Coded year, X	11	12	13	14	15	16	17	18	19	20
Speed, Y	89.6	94.5	91.0	98.2	101.1	95.9	97.5	99.5	97.6	100.4

Year	1931	1932	1933	1934	1935	1936	1937	1938	1939
Coded year, X	21	22	23	24	25	26	27	28	29
Speed, Y	96.6	104.1	104.1	104.9	106.2	109.1	113.6	117.2	115.0

Note that $n = 27$. Calculate $\bar{x} = 15.56$, $\bar{y} = 96.21$, $Sx^2 = 1,908.67$, $Sy^2 = 3,398.60$, $Sxy = 2,454.69$.

EXAMPLE 6.29—Estimate the population mean speed for the year 1917. Ans. 85.2 miles per hour with 95% limits, 83.44 — 86.96 miles per hour.

EXAMPLE 6.30—Forecast the 1941 speed. Ans. 116.07 miles per hour. Remembering that an individual speed is being forecast, not the population mean, calculate the 95% fiducial limits. Ans. 109.16 — 122.98 miles per hour. Actually, the 1941 speed, in the last pre-war race, was 115.12 miles per hour.

EXAMPLE 6.31—On the graph of speed against date, one can draw the curves representing the fiducial limits for each year's speed. These will be closest to the regression line at the mean year (coded $\bar{x} = 15.56$, corresponding to $1910 + 15.56 = 1925.56$), and will deviate further from the regression (measured vertically) as X departs from its mean either way. The curves are the two branches of an hyperbola. If $t_{.05}$ is selected, about 95% of the records should fall between the curves if the sampling is random.

EXAMPLE 6.32—It was desired to forecast the yields of 29 double and three-way crosses of maize under heavy second brood chinch bug infestation (10). The basis of prognostication was the known mean yields of top crosses of the parental lines grown under the same conditions. A formula designed by Jenkins (11) was used. The forecasted yields, X , and the actual experimental yields, Y , read from a scatter diagram, were as follows: 20, 20; 24, 29; 25, 34; 26, 35; 29, 43; 31, 11; 33, 24; 34, 35; 34, 45; 35, 34; 36, 34; 36, 36; 37, 35; 37, 46; 37, 42; 41, 41; 42, 43; 43, 43; 43, 55; 50, 53; 50, 55; 51, 54; 51, 55; 52, 42; 55, 55; 56, 54; 56, 51; 57, 50; 71, 61 bushels per acre. Construct the scatter diagram. Do you think the regression is linear? Assuming that it is, calculate the regression of Y on X , $\hat{Y} = 0.79X + 9.4$.

EXAMPLE 6.33—In the course of his investigations of winter hardiness in apple trees, Stark (14) wished to determine the specific heat, Y , of the twigs. The laboratory technique was difficult and inaccurate. Since more than half the weight of the twigs was water, and since the heat capacity of water is great, he tried estimating Y from the percentage of water, X , in the twigs, the latter being easily and accurately measured. Twenty-one pairs of Stark's measurements, X and Y , are as follows: 49.4, 0.646; 50.1, 0.644; 50.8, 0.665; 51.2, 0.670; 51.5, 0.666; 51.9, 0.653; 52.5, 0.669; 52.7, 0.657; 53.1, 0.689; 53.6, 0.669; 55.7, 0.685; 56.3, 0.696; 57.0, 0.700; 58.0, 0.690; 58.5, 0.711; 59.2, 0.704; 59.7, 0.696; 61.3, 0.713; 62.0, 0.719; 63.1, 0.731; 64.9, 0.731. Compute $\bar{x} = 55.83\%$, $Sx^2 = 431.31$, $\bar{y} = 0.6859$, $Sy^2 = 0.0142$, $Sxy = 2.3536$, $\hat{Y} = 0.005457X + 0.3812$. Since $s_y = 0.027$ and $s_{y \cdot x} = 0.0086$, about a third of the variation of Y is unassociated with X . Since X is determined with very little error, Stark concluded that Y could be determined with sufficient accuracy by first measuring the percentage of water in the apple twigs, then estimating specific heat from the regression equation.

EXAMPLE 6.34—Derive the formula, $Sxy = SXY - (SX)(SY)/n$. Hint: Start with $X = x + \bar{x}$ and $Y = y + \bar{y}$. Multiply and add, remembering that $Sx = Sy = 0$.

6.11—Use of regression in experimental design. Statistical control. As anticipated in section 3.10, regression is one means of reducing the unexplained error encountered in experimentation. Variation not controlled in the conduct of the experiment is often associated with some measurable variate, and the measurement of this variate enables the investigator to use regression for increasing the information available. An example is found in the rat gains of table 6.6. Though it was known that gain is usually associated with initial weight, it was not feasible to control this extraneous variate; that is, to use only rats all having the same starting weight. It was easy, however, to measure the variate, X , and to calculate the regression of gain upon it. What advantage was attained? Without this precaution experimental error would have been based on the rat variance,

$$s^2 = S_y^2/(n - 1) = 6,357.73/14 = 454.12$$

But utilization of known initial weights reduced the error variance to

$$s_{y \cdot x}^2 = 371.91$$

Following the argument of section 3.10, it may be said that the information *per rat* has been increased in the ratio, $454.12/371.91 = 122\%$. Another way of saying it is that, for the same reliability, 22% more rats would have been required if the facts about initial weight had not been utilized. The relatively inexpensive measurement and use of initial weight increased the information from this experiment by 22%.

Even if experimental control is feasible it is not always desirable. Thus, in the experiment with tractor-drawn plows one would certainly not wish to confine his trials to a single speed. On the other hand, he would get little information if he considered

$$s^2 = Sd^2/(n - 1) = 83,240/9 = 9,249$$

as the experimental variance. This experiment having been designed for use of regression, the appropriate estimate of random experimental variance is

$$s_{y \cdot d}^2 = 382.0$$

Not only is the experiment founded on a broader basis of experience, but the information has been increased in the ratio of $9,249/382.0 = 2421\%$; that is, the information per trial is 2321% greater than it would have been if no use had been made of the speed records.

It appears, then, that in many situations where it is either not convenient or not desired to control experimentally a variate associated with the one under investigation, *statistical control* (15) may be introduced. By measuring the auxiliary quantity and using regression one may accomplish the same end as by keeping this quantity constant in the experiment. The error, $s_{y \cdot x}$, is presumably no greater than $s_{y \cdot d}$ would be in an experi-

mentally controlled sample with constant X . Indeed, it seems that in many cases statistical control is more to be desired; the actual situation is studied instead of one artificially produced, the observations are extended over a greater range, thus broadening the foundation for inference, and in the end one has knowledge of the variation of two quantities instead of one, together with the relation between them. In each line of research it must be learned by experience which type of control is more readily available, and which produces information more quickly and cheaply.

EXAMPLE 6.35—The chick data of example 6.7 were adapted from the original records of 13 birds whose body weights (X grams) and comb weights (Y mg.) were as follows; 83, 54; 72, 42; 69, 29; 78, 37; 69, 18; 90, 84; 90, 56; 95, 107; 95, 90; 91, 68; 75, 31; 70, 48; 76, 41. From these records, $n = 13$, $\bar{x} = 81.0$ grams, $\bar{y} = 54.2$ mg., $Sx^2 = 1,218.00$, $Sy^2 = 8,292.31$, $Sxy = 2834.00$. If decisions are to be based on comb weight, estimate the increase in information due to use of body weight as a statistical control. Ans. 348%.

EXAMPLE 6.36—From the apple data of table 6.5, estimate the decrease in precision (information) that would result from omitting crop size as a statistical control. Ans. 75.3%.

EXAMPLE 6.37—For 1594 farms in Crop Estimates District 9 of eastern South Dakota (1942), X = acres in farm and Y = acres in corn were recorded. The variance of corn acres was $s_y^2 = 1,859$, while that of deviations from regression was $s_{y \cdot x}^2 = 1,002$. If you collected a random sample of 100 farms from which to estimate the average number of acres in corn, s_y^2 would be 18.59. If you used farm size as a statistical control, how many farms would be required in the sample to get the same amount of information? Ans. 54 farms.

6.12—Partitioning the sum of squares of the dependent variate. Regression computations may be looked upon as a process of partitioning SY^2 into 3 parts which are both useful and meaningful. You have become accustomed to dividing SY^2 into $(SY)^2/n$ and the remainder, Sy^2 ; then subdividing Sy^2 into $(Sxy)^2/Sx^2$ and $Sd_{y \cdot x}^2$. This means that you have divided SY^2 into three portions:

$$SY^2 = (SY)^2/n + (Sxy)^2/Sx^2 + Sd_{y \cdot x}^2$$

Each of these portions can be associated exactly with the sum of squares of a segment of the ordinates, Y . To illustrate this a simple set of data has been set up in table 6.9 and graphed in figure 6.5.

TABLE 6.9
DATA SET UP TO ILLUSTRATE THE PARTITION OF SY^2

X	2	4	6	8	10	12	14	$SX = 56$
Y	4	2	5	9	3	11	8	$SY = 42$

$$n = 7, \quad \bar{x} = 8, \quad \bar{y} = 6, \quad Sx^2 = 112, \quad Sy^2 = 68, \quad Sxy = 56$$

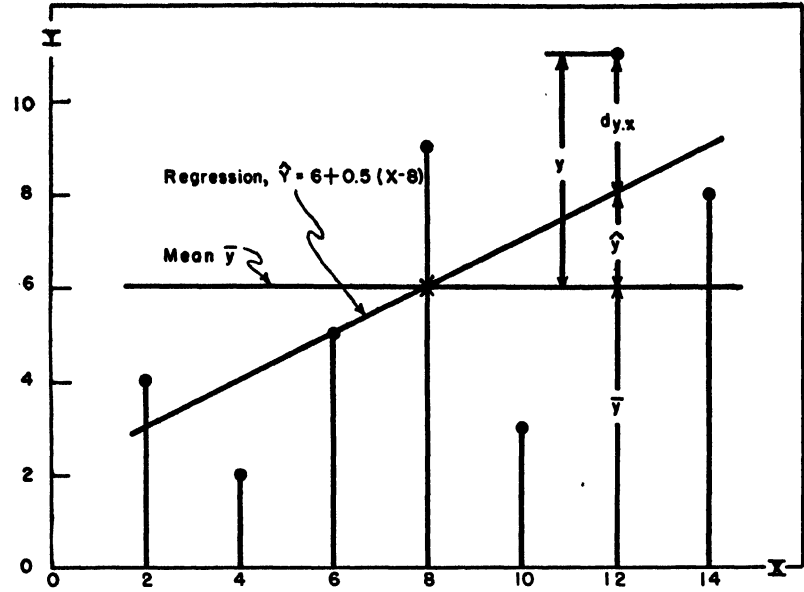


FIG. 6.5—Graph of data in table 6.9. The ordinate at $X = 12$ is shown divided into 2 parts, $\hat{y} = 6$ and $y = 5$. Then y is subdivided into $\bar{y} = 2$ and $d_{y,x} = 3$. Thus $Y = \bar{y} + \hat{y} + d_{y,x} = 6 + 2 + 3 = 11$.

In the figure the sixth ordinate is shown partitioned into 3 segments:

$$Y = \bar{y} + \hat{y} + d_{y,x}$$

Each of the other ordinates may be divided similarly, though negative segments make the geometry less obvious. The lengths are all set out in table 6.10 and the several segments are emphasized in figure 6.6. Observe

TABLE 6.10
LENGTHS OF ORDINATES IN TABLE 6.9 TOGETHER WITH SEGMENTS INTO WHICH THEY ARE PARTITIONED

Pair Number	Ordinate Y	Mean $\bar{y}$	Deviation $\hat{y}$	Deviation From Regression $d_{y,x}$
1	4	6	-3	1
2	2	6	-2	-2
3	5	6	-1	0
4	9	6	0	3
5	3	6	1	-4
6	11	6	2	3
7	8	6	3	-1
	42	42	0	0
Sum of squares	320	252	28	40

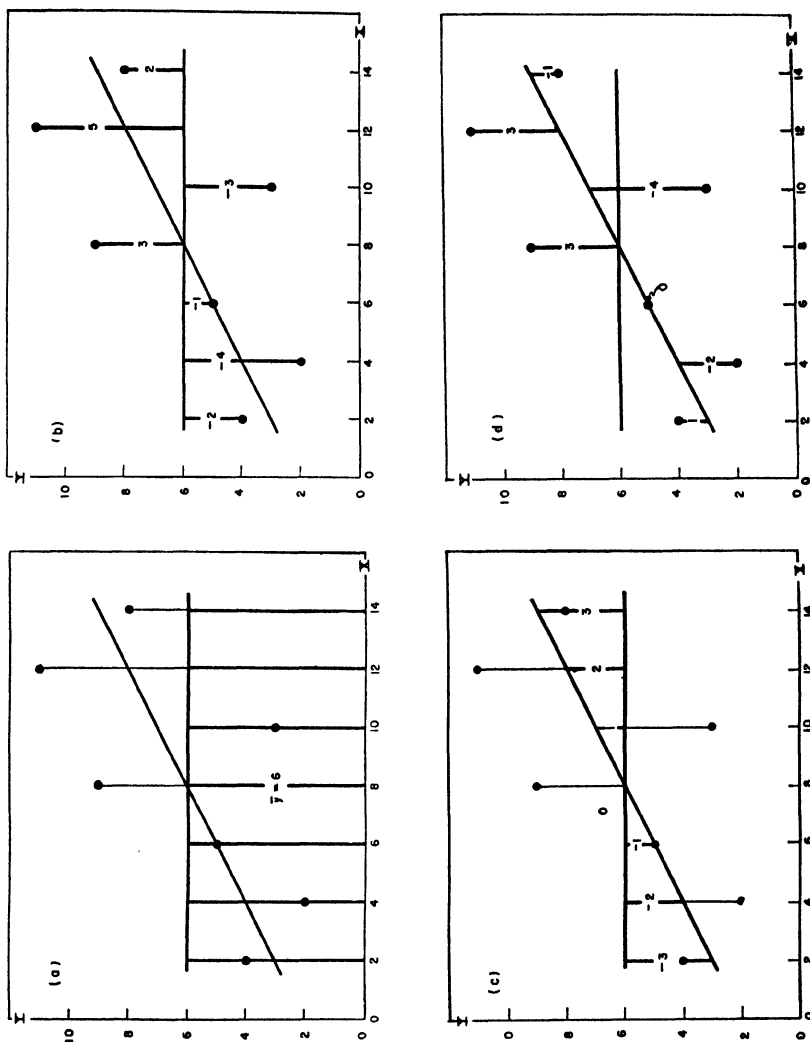


FIG. 6.6(a)—Corresponding to the mean, $\bar{y}$, these are the segments subtracted from Y leaving the deviations from mean, y , shown in figure 6.6(b). The sum of the squares of these 7 segments, $7(6^2) = 252$, corresponds to the 1 degree of freedom for the correction term, $(SY)^2/n = (42)^2/7 = 252$.

FIG. 6.6(b)—These are deviations from mean, y , further subdivided in figures 6.6(c) and (d).

FIG. 6.6(c)—Heavy lines show the deviations of regression values, $\hat{Y}$, from their mean, $\bar{y}$. They are completely determined by $\bar{y}$, b and a fixed set of x 's. The sum of their squares, $SY^2 = 28$, is the portion of SY^2 attributable to regression.

FIG. 6.6(d)—Random deviations from regression, $d_{y.x}$. Each $d_{y.x}$ added to the corresponding $\hat{y}$ and $\bar{y}$ of figures 6.6(a) and (c) comprises an ordinate, Y , in figure 6.5. The sum of the squares of these deviations, $Sd_{y.x}^2 = 40$, corresponding to 5 degrees of freedom, furnishes an estimate, $s_{y.x}^2 = 40/5 = 8$, of σ^2 .

that in each line of the table (including the two at the bottom) the sum of the last three numbers is equal to the number in column Y .

It is notable that, though

$$Y = \bar{y} + \hat{y} + d_{y \cdot x},$$

yet the sum of squares is

$$SY^2 = S\bar{y}^2 + S\hat{y}^2 + Sd_{y \cdot x}^2,$$

each of the three product terms being zero (example 6.42). The sums of squares of the ordinates, $SY^2 = 320$, and of the deviations from regression, $Sd_{y \cdot x}^2 = 40$, are already familiar. It remains only to identify $(SY)^2/n$ with $S\bar{y}^2$ and $(Sxy)^2/Sx^2$ with $S\hat{y}^2$.

First,

$$(SY)^2/n = n \cdot \frac{(SY)^2}{n^2} = n\bar{y}^2 = S\bar{y}^2$$

That is, the correction for mean is simply the sum of the squares of the mean taken n times. The reason it is not so used in calculation is that any rounding inaccuracy in $\bar{y}$ would be exaggerated both by squaring and by multiplication with n .

Second,

$$\frac{(Sxy)^2}{Sx^2} = \frac{(Sxy)^2}{(Sx^2)^2} Sx^2 = b^2 Sx^2 = Sb^2 x^2 = S\hat{y}^2$$

So the sum of squares attributable to regression turns out to be the sum of the squares of the n deviations of $\hat{Y}$ from its mean, $\bar{y}$. Certainly, it would be a tedious job to calculate $S\hat{y}^2$ every time one wished the sum of squares due to regression. You are now in a position to evaluate the short cut, $(Sxy)^2/Sx^2$.

I hope you appreciate the beauty of these regression computations. Make each formula meaningful, not merely a jumble of letters and figures. Associate each with a column in the table and with a line segment in the figure as well as with the descriptive phrase which has been used. Familiarity breeds not only utility but admiration.

The partition of SY^2 is intimately associated with the division of the total available degrees of freedom into 3 parts, one part corresponding to each portion of SY^2 (table 6.11). This identification of the several sums of squares with corresponding degrees of freedom is one of the useful contributions of Professor R. A. Fisher (7).

It will be observed that zero is the sum of the deviations from the horizontal line of figure 6.6 (a) and (b) as well as from the regression line of figure 6.6 (d). In fact, this would be true of any line through the origin of deviations, $(\bar{x}, \bar{y})$. The sum of the squares of the deviations in figure

TABLE 6.11
PARTITION OF DEGREES OF FREEDOM CORRESPONDING TO 3 SUMS OF SQUARES

Description	Symbol	Degrees of Freedom	Sum of Squares	Variance
The mean	$\bar{y}$	1	$S\bar{y}^2 = (SY)^2/n$	
Deviations determined by b and a fixed set of x 's	$\hat{y} = bx$	1	$S\hat{y}^2 = (S_{xy})^2/Sx^2$	
Random deviations from regression	$d_{y \cdot x}$	$n - 2$	$Sd_{y \cdot x}^2$	$s_{y \cdot x}^2$
Total	Y	n	SY^2	

6.6 (b) is less than the corresponding sum from any other *horizontal* line. But if you begin to rotate this line counterclockwise about the center, $(\bar{x}, \bar{y})$, the sum of squares becomes smaller and smaller till the position of the regression line is reached. It is the introduction of a second variate, such as crop size in the apple data, that gives meaning to this rotation. So long as there is only injured apples to consider, with no hint as to differences in crop size, the horizontal line is the only one interpretable. In fact, all the points might as well be plotted on some vertical line, such as OY of figure 6.3. In that event, the deviations from the origin, $\bar{y}$, would be exactly those considered in the earlier chapters. But added information about crop size warrants the extension of our ideas to include regression. Much of the variation in percentage injured fruits is eventually explained by differences in crop size. The only remaining variation is that asso-

EXAMPLE 6.38—Dawes (4) determined the "density" of the melanin content of the skin of 24 male frogs together with their weights. Since "Some of the 24 males . . . were selected for extreme duskiness or pallor so as to provide a measure of the extent of variability," that is, since selection was exercised on density this variate cannot be estimated (section 6.3) from the sample. Assuming random weights, they may be taken as Y . The data follow:

Density, X Weight, Y	0.13 13	0.15 18	0.28 18	0.58 18	0.68 18	0.31 19	0.35 21	0.58 22
Density, X Weight, Y	0.03 22	0.69 24	0.38 25	0.54 25	1.00 25	0.73 27	0.77 27	0.82 27
Density, X Weight, Y	1.29 28	0.70 29	0.38 30	0.54 30	1.08 35	0.86 37	0.40 39	1.67 42

Calculate $\bar{x} = 0.6225$ units, $\bar{y} = 25.79$ grams, $Sx^2 = 3.3276$, $Sy^2 = 1,211.96$, $Sxy = 40.022$.

ciated with the deviations from the regression line due, according to our hypothesis, to sampling variation of injury to the apple crop. The selection of that line for which the sum of such deviations is zero, and the sum of their squares a minimum, is an effect of the logical extension of those ideas that led to the selection of the mean as an appropriate average of concentration. The principle of least squares is simply a statement of the unity in our ideas of averages for single variates and regressions for two or more.

EXAMPLE 6.39—In example 6.38 test the hypothesis, $\beta = 0$. Ans. $t = 3.81$, $P < 0.01$.

EXAMPLE 6.40—In the frog data divide SY^2 into its 3 components, thus:

Source of Variation	Degrees of Freedom	Sum of Squares	Variance
Mean	1	15,965.04	
Regression	1	481.36	
Deviations	22	730.60	33.21
Total	24	17,177.00	

EXAMPLE 6.41—How nearly free from error is the measurement of melanin density, X ? After preparation of a solution from the skin of the frogs, the intensity of the color was evaluated in a colorimeter and the readings then transferred graphically into neutral densities. The figures reported are means of from 3 to 6 determinations. The error of this kind of measurement is usually appreciable, so that the estimate of regression may be biased downwards. Had not the investigator wished to learn about extremes of density, the regression of density on weight might have been not only unbiased but more informative.

EXAMPLE 6.42—Prove that the sums of the cross products in SY^2 are all zero. The only one that offers any difficulty is $SYd_{y..}$: set $\hat{y} = bx$ and $d_{y..} = y - \hat{y} = y - bx$. Multiply and add, noting that $Sxy = bSx^2$.

6.13—Galton's use of the term "regression." In his studies of inheritance Galton developed the idea of regression. Of the "law of universal regression" (8) he said, "Each peculiarity in a man is shared by his kinsman, but on the average in a less degree." His friend, Karl Pearson (12), collected more than a thousand records of stature, cubit and span in family groups. Figure 6.7 shows his regression of son's stature on father's. Though tall fathers do tend to have tall sons, yet the average height of sons of a group of tall fathers is less than their father's height. There is a regression, or going back, of sons' heights toward the average height of all men.

As part of Galton's evidence he reported the results of an experiment on the inheritance of size in sweet pea seeds. For planting he selected 100 seeds in each of seven size groups, then ascertained the average size of the offspring. He gives the following data on diameter of parent sweet peas and mean diameter of filial seeds: 21, 17.5; 20, 17.3; 19, 16.0;

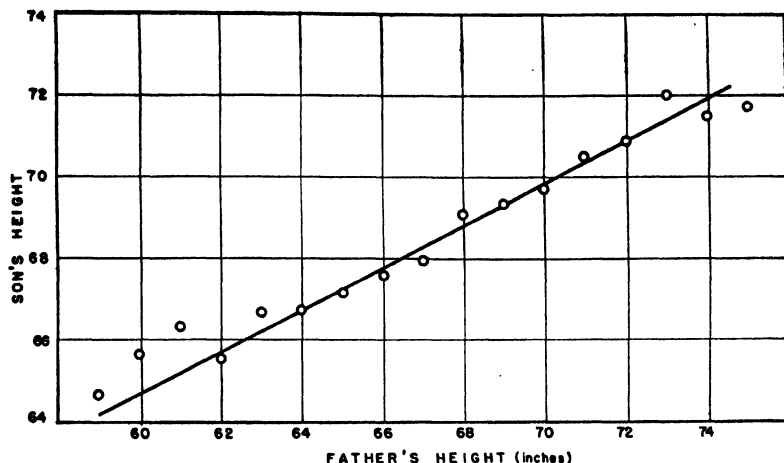


FIG. 6.7—Regression of son's stature on father's (12). $\hat{Y} = 0.516X + 33.73$. 1078 families.

18, 16.3; 17, 15.6; 16, 16.0; 15, 15.3; all measurements in hundredths of an inch.

The regression coefficient of filial size on parent size, that is Y on X , was 0.3. Galton thought that regression was generally about one-third, the average of the offspring deviating from the racial mean only about a third as much as the deviation of the same measurement in the lot of parents having a common measure.

If you are a student of genetics you will realize that this is not a simple phenomenon, involving as it does such concepts as heterozygosity, dominance, etc. An additional complexity in stature is the tendency toward the *selective mating* of men and women of similar stature. Apparently this law of universal regression summarizes those tendencies in a species to preserve its unique characteristics. If tall fathers begot taller daughters and tall daughters bore taller sons there would be no regression, and the two related concepts, *species* and *average*, would presumably not exist. The average of concentration in statistics is that measurement of a character toward which regression is directed, while the average of variation stays constant from generation to generation because of this regression.

EXAMPLE 6.43—Warren (16) investigated the magnitude of regression in parthenogenetic generations of aphids. He measured the frontal breadth in units of 0.0221 mm., the results for parents (X) and offspring (Y) being as follows: 14.8, 13.7; 18.4, 14.6; 17.5, 14.5; 16.3, 14.3; 16.9, 14.7; 15.8, 14.2; 15.4, 14.8; 15.2, 15.0; 15.9, 15.8; 16.1, 15.7; 16.3, 15.1; 16.7, 15.6; 17.0, 15.5; 17.8, 15.9; 18.1, 15.4; 18.5, 15.8; 18.9, 15.2; 15.8, 16.2; 16.0, 16.0; 16.2, 16.7; 16.6, 16.6; 16.8, 16.9; 17.5, 16.4; 18.2, 16.3; 18.8, 16.7; 19.0, 16.9; 15.7, 17.2; 16.1, 17.0; 16.5, 17.6; 16.8, 17.4; 17.1, 17.8; 17.7, 17.5; 18.0, 17.7; 18.8, 17.9; 19.5, 17.6; 16.0, 18.1; 17.5, 18.8; 19.0, 18.4; 18.9, 19.5. Compute the regression coefficient, 0.4469.

6.14—Regression and rate (or ratio). In section 6.2 you learned that the growth rates, Y/X , could be appropriately averaged if two conditions were met: (i) the regression passes through the origin, and (ii) the weekly standard deviation increases in proportion to age. An example of this kind was encountered in studying the percentage of farm land in corn among eastern South Dakota farms (1942). For illustration we selected 5 common farm sizes and chose at random 20 farms from each (table 6.12).

TABLE 6.12
DATA FROM EASTERN SOUTH DAKOTA FARMS, 1942
Twenty farms chosen at random from each selected size group.

Size of Farm (acres) X	Average Number Acres in Corn Y	Standard Deviation (acres) s_y	Ratio s_y/X	Ratio Y/X
80	21.30	8.50	0.106	0.2662
160	42.00	15.98	0.100	0.2625
240	66.25	20.38	0.085	0.2760
320	64.60	28.51	0.089	0.2019
400	88.90	39.90	0.100	0.2222
Mean	240	56.61		0.2458

The ratio, s_y/X , is so nearly constant that the population standard deviation may be assumed proportional to acreage. If the further reasonable assumption is made, that the population regression contains the origin, then the least squares estimate of the change in corn acres per acre increase in farm size is the mean of the ratios, Y/X ; that is, 0.2458 or 24.58%. The regression equation,

$$\hat{Y} = 0.2458X,$$

leads to the estimate of corn acreage in the average sized farm (240 acres),

$$\hat{Y} = (0.2458)(240) = 58.99 \text{ acres}$$

If you plot corn acres, Y , against farm acres, X , then plot the regression, you will see the effect of the small weights assigned to the larger farm sizes on account of their larger standard deviations (lesser reliability). The regression line hugs the dots for the smaller farms. Another point of interest is this: the average number of corn acres calculated in column 2 of the table is not the correct least squares estimate, 58.99 corn acres for a 240-acre farm, but is biased downwards. In other words, the point $A(\bar{x}, \bar{y})$ is not on this regression line fitted under the two conditions specified above.

In view of the extensive use of rates and ratios, it may be well to examine a situation in which the practice is questionable. Employing appropriate regression methods, Crampton (3) reported the gains in weight (Y) of pigs eating recorded amounts of feed (X). For his lot I (table 6.13),

TABLE 6.13
DATA ON LOT I OF CRAMPTON'S PIGS

	Feed Eaten (pounds) X	Gain (pounds) Y	Ratio Y/X	Predicted Gain	
				$\hat{Y} = 0.2627X - 22.1$	$\hat{Y} = 0.1999X$
	382	66	0.1728	78.3	76.4
	335	72	0.2149	65.9	67.0
	388	84	0.2165	79.8	77.6
	316	47	0.1487	60.9	63.2
	319	75	0.2351	61.7	63.8
	399	87	0.2180	82.7	79.8
	358	75	0.2095	72.0	71.6
	355	73	0.2056	71.2	71.0
	344	59	0.1715	68.3	68.8
	339	70	0.2065	67.0	67.8
Mean	353.5	70.8	0.1999	70.8	70.7

$n = 10$, $\bar{x} = 353.5$ pounds of feed, $\bar{y} = 70.8$ pounds gain, $Sx^2 = 7,334.5$, $Sy^2 = 1,207.5$, $Sxy = 1,927.0$. Figure 6.8 displays the scatter diagram

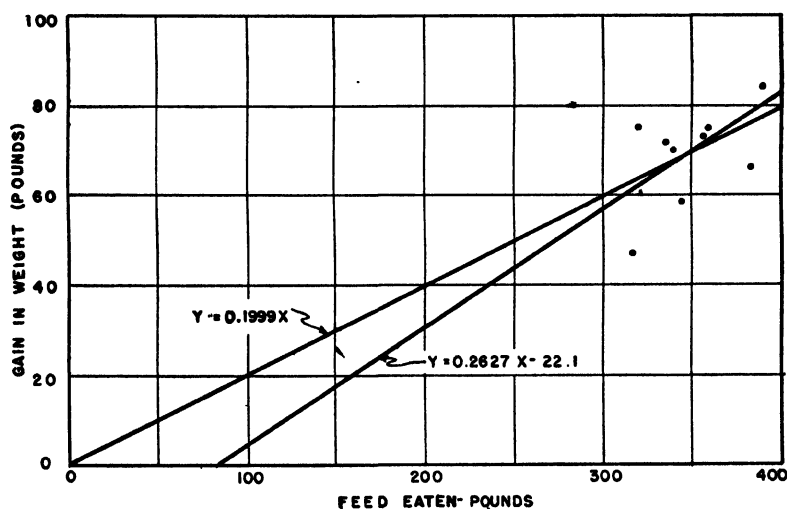


FIG. 6.8—Graph of regressions of gain on food consumed. Crampton's pigs, lot I.

with his regression, $\hat{Y} = 0.2627X - 22.1$. Crampton called attention to the fact that this regression does not contain the origin because feed is required to maintain the animal even with no increase in weight. Furthermore, the sample evidence admits the hypothesis of constant variance in Y over the range found, and there is no appreciable error in the measure-

ment of X . It would seem correct, therefore, to employ the common method used throughout this chapter beginning with section 6.5.

In this rather common situation, if the ratios, Y/X (feed required per pound gain), were used, they would lead to inefficient estimates. The regression coefficient in column 3 of table 6.13, 0.1999 pound gain per pound of feed (or 19.99 pounds gain per 100 pounds of feed) underestimates the gains of pigs eating heavily and overestimates the others: the regression line through the origin does not depict the behavior of the animals. The variance from regression through the origin is 91.28 instead of the efficient estimate, 87.66, indicating a loss of 4% from the information available in the data.

In the regression, $\hat{Y} = 0.2627X - 22.1$, the ratio of the deviations, $\hat{y}/x = 0.2627$, is constant; but the ratio, $\hat{Y}/X$, increases with X . This is clear from dividing both members of the regression equation by X :

$$\frac{\hat{Y}}{X} = 0.2627 - \frac{22.1}{X}$$

For a pig eating 300 pounds of feed, $\hat{Y}/X = 0.2627 - 22.1/300 = 0.1890$ pound gain per pound of feed, while for one eating 400 pounds, $\hat{Y}/X = 0.2075$. This means that the rates, Y/X , for two animals are not directly comparable unless X is the same in both. Comparison could be made by calculating the expected rates, $\hat{Y}/X$, then contrasting the deviations of actual rates from expected. The better way is by use of adjusted Y or else deviations from regression (section 6.7).

Unless there is reason to believe that the regression passes through the origin, the ratio, Y/X , is of dubious value. It may do for rough work, but careful experimentation deserves the more efficient statistical method.

6.15—Summary. If, in sampling from a bivariate population, Y is taken at random for either random or selected values of X measured without error, the regression coefficient is an unbiased estimate of β , the change in Y per unit change in X . If X is subject to error, b is biased, but not seriously unless the error in X is large as compared to the range. Estimates of β are given for homogeneous variance in Y , and for these additional cases if the regression contains the origin: (i) standard deviation increases with X , and (ii) variance increases with X .

The t -test of null hypotheses about β is explained. Fiducial limits are calculated for $\hat{y}$, b , $\hat{Y}$, and a random Y for any X . Methods are described for case studies and estimates of differences between individuals and group means after adjustment to a common X .

The role of regression and statistical control in experimental design is discussed. SY^2 together with the associated n degrees of freedom is partitioned into $(SY)^2/n$ (1 degree of freedom), $(Sxy)^2/Sx^2$ (1 degree of freedom) and $Sd_{y..}^2$ ($n - 2$ degrees of freedom).

Galton's use of the term, regression, is illustrated. There is a discussion of the inefficient use of the rate (or ratio), Y/X .

With the six quantities, n , $\bar{x}$, $\bar{y}$, Sx^2 , Sy^2 , and Sxy , at hand, the following are the principal formulas used:

1. $b = Sxy/Sx^2$, the regression coefficient, Y on X .
2. $\hat{Y} = bX$ or $\hat{Y} = a + bX$, the regression equation, Y on X .
3. $Y - bx$, the adjusted value of Y .
4. $Y = \bar{y} + \hat{y} + d_{y \cdot x}$, the mean of y plus the deviation of $\hat{Y}$ from its mean, $\bar{y}$, plus the deviation from regression, $Y - \hat{Y}$.
5. $SY^2 = S\hat{y}^2 + S\hat{y}^2 + Sd_{y \cdot x}^2$, the partition of SY^2 into three parts.
6. $(Sxy)^2/Sx^2 = S\hat{y}^2$, the sum of squares attributable to regression.
7. $Sy^2 - (Sxy)^2/Sx^2 = Sd_{y \cdot x}^2$, the sum of squares of deviations from regression.
8. $s_{y \cdot x}^2 = Sd_{y \cdot x}^2/(n - 2)$, the variance from regression or variance of estimate.
9. $s_{y \cdot x}$, the standard deviation from regression or standard error of estimate.
10. $s_{\hat{y}} = s_{y \cdot x}/\sqrt{n}$ for a fixed set of X 's.
11. $s_b = s_{y \cdot x}/\sqrt{Sx^2}$, the standard error of b .
12. $s_{y \cdot x}\sqrt{1/n + x^2/Sx^2}$, the standard error of $\hat{Y}$.
13. $s_{y \cdot x}\sqrt{1 + 1/n + x^2/Sx^2}$, the standard error of an estimated Y .

REFERENCES

1. A. E. BRANDT. *Journal of the American Statistical Association*, 23:291 (1928).
2. E. V. COLLINS. *Transactions, American Society of Agricultural Engineers*, 14:39 (1920).
3. EARLE WILCOX CRAMPTON. *Journal of Nutrition*, 7:305 (1934).
4. BEN DAWES. *Journal of Experimental Biology*, 18:26 (1941).
5. W. EDWARDS DEMING. *Statistical Adjustment of Data*. John Wiley and Sons, Inc., New York (1943).
6. GEORGE H. DUNGAN. *Plant Physiology*, 9:749 (1934).
7. R. A. FISHER. *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh (1925-1944).
8. FRANCES GALTON. *Natural Inheritance*. Macmillan and Company, London (1889).
9. T. ROY HANSBERRY and CHARLES H. RICHARDSON. *Iowa State College Journal of Science*, 9:414 (1935).
10. J. R. HOLBERT, W. P. FLINT, J. H. BIGGER, and G. H. DUNGAN. *Iowa State College Journal of Science*, 9:414 (1935).
11. M. T. JENKINS. *Journal of the American Society of Agronomy*, (1934).
12. KARL PEARSON and ALICE LEE. *Biometrika*, 2:357 (1903).
13. GEORGE W. SNEDECOR and W. R. BRENNEMAN. *Iowa State College Journal of Science*, 19:333 (1945).
14. ARVIL L. STARK. *Plant Physiology*, 8:168 (1933).
15. H. A. WALLACE and GEORGE W. SNEDECOR. *Correlation and Machine Calculation*. Iowa State College Official Publication 30, No. 4 (1931).
16. ERNEST WARREN. *Biometrika* 1:129 (1902).
17. J. B. WENTZ and R. T. STEWART. *Journal of the American Society of Agronomy*, 16:534 (1924).

Chapter 7

CORRELATION

7.1—Introduction. Bivariate populations are often interesting, not because one quantity is to be predicted from the other, but because of the mutual relationship between the measurements. Some evaluation of this relation is needed even if there is no reason for thinking of one variate as dependent on another. Galton's proposal of this problem sprang from his study of inherited characteristics, a study attaining scientific rank during the latter half of the nineteenth century. Biologists and mathematicians vied with each other in developing the theory and its applications. It turned out that the subject of *correlation*, as it came to be called, acquired a group of the concepts of regression, so that in many writings the two are scarcely distinguished.

Correlations among stature, span and cubit (length of forearm) were reported by Pearson and Lee (18) who collected data from more than a thousand British families. The measurements of brother and sister heights in table 7.1 are selected to illustrate one of their tables. Since there is

TABLE 7.1
STATURE (INCHES) OF BROTHER AND SISTER
Illustration taken from Pearson and Lee's sample of 1,401 families.

Family number	1	2	3	4	5	6	7	8	9	10	11
Brother, X_1	71	68	66	67	70	71	70	73	72	65	66
Sister, X_2	69	64	65	63	65	62	65	64	66	59	62

$$n = 11 \quad \bar{x}_1 = 69, \quad \bar{x}_2 = 64, \quad Sx_1^2 = 74, \quad Sx_2^2 = 66, \quad Sx_1x_2 = 39$$

$$r = Sx_1x_2 / \sqrt{(Sx_1^2)(Sx_2^2)} = 39 / \sqrt{(74)(66)} = 0.558. \quad \text{Pearson and Lee's } r = 0.553$$

here no distinction between the dependence of the two variates, they are designated merely as X_1 and X_2 instead of X and Y . You will have no difficulty in verifying the calculation of the sums of squares and products. The new feature is the *correlation coefficient*, universally symbolized by r . It is a numeralization of that commonly observed similarity among children of the same parents, the tendency of the taller sisters to have the taller brothers. This is represented graphically by a diagram such as figure 7.1. The value, $r = 0.558$, reflects the propensity of the dots to lie in a band

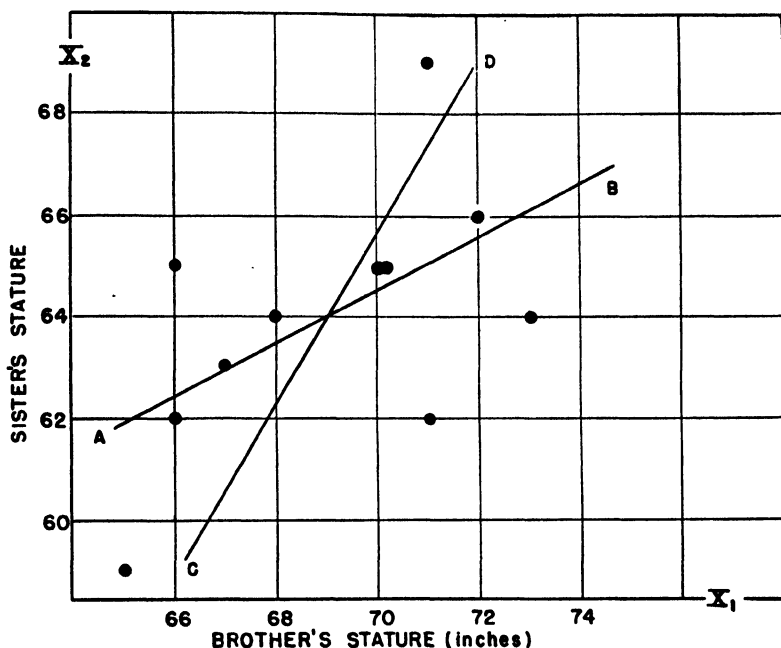


FIG. 7.1—Scatter (or dot) diagram of stature of 11 brother-sister pairs. $r = 0.558$

extending from lower left to upper right, not scattered randomly over the whole field. These points are confined to an elliptical area with the major axis inclined toward the right. Biologically, $r = 0.558$ measures the common inheritance of stature by these siblings.

EXAMPLE 7.1—Calculate $r = 1$ for the following pairs:

X_1 : 1, 2, 3, 4, 5

X_2 : 3, 5, 7, 9, 11

Represent the data in a graph similar to figure 7.1.

EXAMPLE 7.2—Verify $r = 0.91$ in the pairs:

X_1 : 2, 5, 6, 8, 10, 12, 14, 15, 18, 20

X_2 : 1, 2, 2, 3, 2, 4, 3, 4, 4, 5

Plot the elliptical band of points.

EXAMPLE 7.3—In the following, show that $r = 0.20$:

X_1 : 3, 5, 8, 11, 12, 12, 17

X_2 : 11, 5, 6, 8, 7, 18, 9

Observe the scatter of the points in a diagram.

EXAMPLE 7.4—In the apple data of table 6.5, $S_x^2 = 924$, $S_y^2 = 1,222$, $S_{xy} = -936$. Calculate $r = -0.88$.

7.2—The correlation coefficient, r . The expression, $S_{x_1x_2}/\sqrt{(S_{x_1}^2)(S_{x_2}^2)}$, is designed to vary between -1 and $+1$ according to the closeness of the

relationship in the population sampled. Negative values of r , like those of b , indicate an inclination of the ellipse of points downward toward the right, large values of one variate being associated with small values of the other. To help you acquire some experience of the nature of r , a number of simple tables with the corresponding graphs are displayed in figure 7.2.

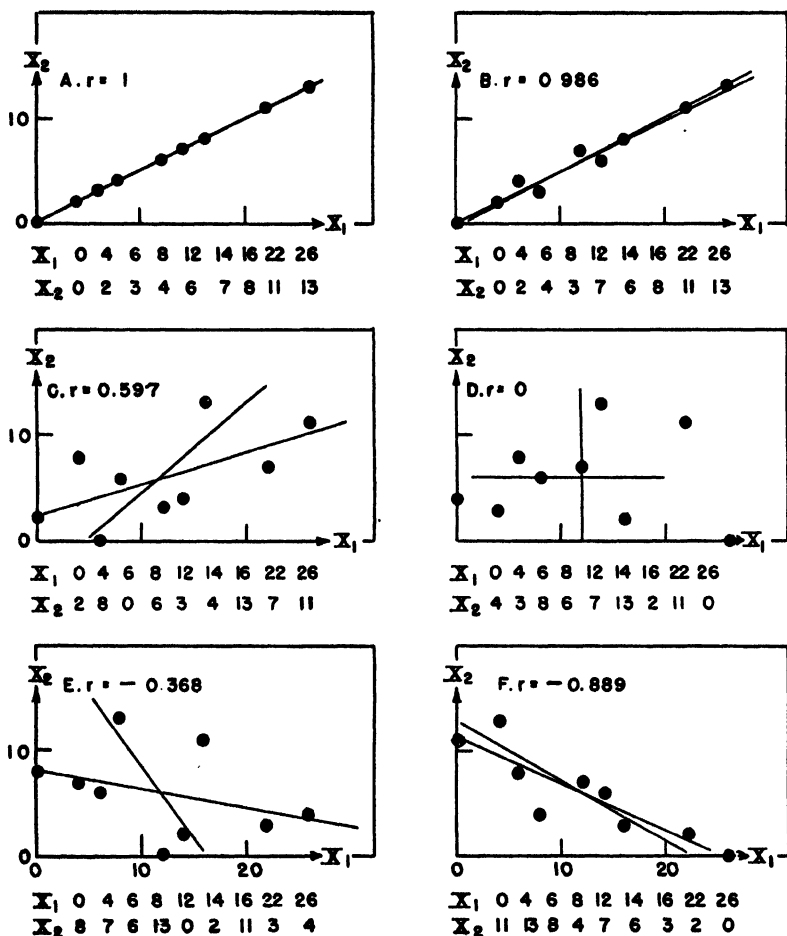


FIG. 7.2—Scatter diagrams with correlations ranging from 1 to -0.889 .

In each of these tables $n = 9$, $\bar{x}_1 = 12$, $\bar{x}_2 = 6$, $Sx_1^2 = 576$, $Sx_2^2 = 144$. Only Sx_1x_2 changes, and with it the value of r . Since $\sqrt{(Sx_1^2)(Sx_2^2)} = \sqrt{(576)(144)} = 288$, the correlation is easily evaluated in the several tables by calculating Sx_1x_2 and dividing by 288 (or multiplying by $1/288 = 0.0034722 \dots$ if a machine is used).

In *A*, the nine points lie on a straight line, the condition for $r = 1$. The line is a "degenerate" ellipse—it has length but no width. In this relation between the two variates they keep in perfect step, any change in one being accompanied by a proportionate change in the other. *B* depicts some deviation from linearity, the ellipse being long and thin with r slightly reduced below 1. In *C*, the ellipse widens, then reaches circularity in *D* where $r = 0$. This denotes no relation between the two variates. *E* and *D* show negative correlations tending towards -1 . Summarizing: the thinness of the ellipse of points exhibits the magnitude of r , while the inclination of the axis to the right or left shows its sign. It should be observed that the slope of the axis is determined by the scales of measurement adopted for the two axes of the graph and is therefore not a reliable indicator of the magnitude of r . It is the concentration of the points near the axis of the ellipse that signifies high correlation, especially if the regression is large.

I think you will agree that the larger correlations, either positive or negative, are fairly obvious from the graphs. It is not so easy to make a visual evaluation if the absolute value of r is less than 0.5; even the direction of inclination of the ellipse may elude you if r is between -0.3 and $+0.3$. In these small samples a single dot can make a lot of difference. In *D*, for example, if the point (26, 0) were changed to (26, 9) the correlation would be increased to 0.505. This emphasizes the fact that sample r 's from a bivariate population in which the correlation is ρ (the Greek rho distinguishes the parameter from the statistic) are quite variable if n is small. The sampling distribution of r will be considered in section 7.4. In assessing the value of r in a table, select some extreme values of one variate and note whether they are associated with extreme values of the other. If no such tendency can be detected, r is likely small.

Perfect correlation ($r = 1$) rarely occurs in biological data, though values as high as 0.99 are not unheard of. Each field of investigation has its own range of coefficients. Characteristics inherited like height ordinarily have correlations between 0.35 and 0.55. Among high school grades r averages around 0.35 (3). Pearson and Lee got "organic correlations," that is, correlations between two such measurements as stature and span *in the same person*, ranging from 0.60 to 0.83. Brandt (2) calculated the correlation, 0.986, between live weight and warm dressed weight of 533 swine. Evvard, *et al.* (6) estimated $r = -0.68$ between average daily gain of swine and feed required per pound gained.

It is clear that judgment about the size of a correlation should be made in the light of similar correlations encountered in the same field, sometimes with little reference to the theoretical limits, ± 1 .

7.3—Calculation of r with coding. If the data have been coded (section 5.4), the correlation coefficient may be computed directly from the coded sums of squares and products. Unlike the regression coefficient, r does not require the decoding of the calculated values. For example,

in table 6.8, $r = 735.2/\sqrt{(674.1)(832.4)} = 0.98$. The same result would attend the use of the decoded values, $r = 1,470.4/\sqrt{(26.96)(83,240)}$.

EXAMPLE 7.5—Thirty students scored as follows in two mathematics achievement tests:

I	73	41	83	71	39	60	51	41	85	88	44	71	52	74	50
II	29	24	34	27	24	26	35	18	33	39	27	35	25	29	13

42	85	53	85	44	66	60	33	43	76	51	57	35	40	76
13	40	23	40	22	25	21	26	19	29	25	19	17	17	35

Calculate $r = 0.774$.

EXAMPLE 7.6—Grout (10) measured the wing length and tongue length, both in millimeters, of 44 bees:

Wing	9.68	9.81	9.59	9.68	9.84	9.59	9.61	9.55	9.25	9.08	9.70
Tongue	6.53	6.71	6.70	6.69	6.70	6.62	6.59	6.55	6.35	6.25	6.61

9.60	9.50	9.74	9.72	9.64	9.73	9.77	9.72	9.54	9.65	9.74
6.51	6.55	6.74	6.75	6.45	6.75	6.70	6.65	6.68	6.77	6.44

9.59	9.71	9.56	9.61	9.61	9.55	9.78	9.74	9.48	9.71	9.20
6.54	6.64	6.55	6.57	6.61	6.64	6.64	6.63	6.62	6.55	6.22

9.53	9.74	9.67	9.56	9.49	9.64	9.45	9.52	9.58	9.60	9.68
6.43	6.67	6.68	6.62	6.71	6.70	6.50	6.41	6.50	6.62	6.69

Code by subtracting 9 from each wing length and 6 from each tongue length. Calculate $r_1 = 0.731$.

EXAMPLE 7.7—Calculate $r = 0.96$ between percentage reduction in leaf area and reduction in yield of shelled corn, example 6.4. Code the first set by dividing each item by 1636. Don't forget the pair (0, 0). Contrast the selection of percentage reduction in leaf area with the specifications of section 7.6.

EXAMPLE 7.8—After reviewing examples 17, 18, and 19 of chapter 5, prove that if $X' = a + bX$ and $Y' = c + dY$, then $r_{a'Y'} = r_{aY}$.

7.4—Relation between the coefficients of correlation and regression. If X_2 were designated as dependent, its regression on X_1 would be $b_{21} = S_{X_1X_2}/S_{X_1}^2$, but if X_1 were dependent, the regression of X_1 on X_2 would become $b_{12} = S_{X_1X_2}/S_{X_2}^2$. These expressions differ in their denominators only. The regression lines corresponding to the two circumstances are drawn in the several diagrams of figure 7.2. If r is near ± 1 the lines are close together, becoming identical in A ; if $r = 0$, they are perpendicular. Note: if X_2 is the independent variate, the slope, b_{12} , is measured from

the X_2 axis, vertical in the figures. The regression of X_1 on X_2 is always the line that makes the lesser angle with the vertical axis.

Now, it is easily seen that

$$b_{21}b_{12} = \frac{Sx_1x_2}{Sx_1^2} \cdot \frac{Sx_1x_2}{Sx_2^2} = \frac{(Sx_1x_2)^2}{Sx_1^2 Sx_2^2} = r^2$$

Thus,

$$r = \sqrt{b_{21}b_{12}}$$

In other words, r is the *geometric mean* of the two regression coefficients that might be calculated: it is intermediate in value between them. Let us illustrate.

If we were interested in estimating the statures of sisters from those of their brothers, we should have (table 7.1),

$$b_{21} = 39/74 = 0.527 \text{ inch sister's height per inch brother's,}$$

graphed as AB in figure 7.1. On the other hand, if the interest were in estimating brothers' heights,

$$b_{12} = 39/66 = 0.591 \text{ inch brother's height per inch sister's,}$$

the regression line being CD . Usually, however, it is the mutual relationship that attracts attention because this is taken to represent the heritability of stature, so we calculate

$$r = \sqrt{(0.527)(0.591)} = 0.558,$$

an average of the two regression coefficients.

This serves to clarify the relation of the two coefficients, correlation and regression, in measuring relationship. The latter is the appropriate one if one variate, Y , may be designated as dependent upon the other, X . Y may partly be controlled or caused by X , as when the available amounts of some glandular secretion cause differences in the sizes of organisms. Or, Y may be subsequent to X , as weight gain in nutrition experiments follows the measurement of initial weight. In such cases, the regression of Y on X is usually the statistic that furnishes the information desired. It is then appropriate to attempt to estimate the value of Y from a knowledge of the corresponding value of X . Correlation, on the other hand, is the appropriate measure of the relation between two variates like statures of sister and brother. The two heights are known to be associated through the complex mechanism of inheritance, but neither may be looked upon as a consequence of the other. In this sense correlation is a two-way average of relationship, while regression is directional. Of course, there are many variates whose relation may be studied by means of either correlation or regression, or both. It is necessary only to keep clearly in mind the character of the relation being considered.

A characteristic of r is revealed by dividing both numerator and denominator of the ratio by $n - 1$:

$$r = \frac{S_{x_1x_2}}{\sqrt{(S_{x_1^2})(S_{x_2^2})}} = \frac{\frac{S_{x_1x_2}}{n-1}}{\sqrt{\frac{S_{x_1^2}}{n-1} \cdot \frac{S_{x_2^2}}{n-1}}} = \frac{\text{covariance}}{\sqrt{(s_1^2)(s_2^2)}} \\ = \frac{\text{covariance}}{\text{geometric mean of variances}}$$

Correlation is thus seen to be the quotient of two averages of variation: one, the covariance of the two measurements, X_1 and X_2 ; the other, an average of the two variances. It is an abstract number measuring *covariation*: if of two related characters each occurs in various sizes, their correlation is a measure of the extent to which their variations are concomitant.

Similarly, b takes on new meaning from the changes below:

$$b_{12} = \frac{S_{xy}}{S_{x^2}} = \frac{S_{xy}}{\sqrt{S_{x^2}S_{y^2}}} \cdot \frac{\sqrt{S_{y^2}}}{\sqrt{S_{x^2}}} = r \frac{\sqrt{S_{y^2}/(n-1)}}{\sqrt{S_{x^2}/(n-1)}} = r \frac{s_y}{s_x}$$

In this manner b is intimately related to r . Either can be calculated from the other if the standard deviations of the two variates are known.

It is often convenient to use the standard deviation as the unit for measuring the deviate, $x = X - \bar{x}$. The *standard deviate*, $x' = x/s$, has the average range given in table 5.5 for $\sigma = 1$; that is, of course, if samples are randomly drawn from normal populations. As an example, suppose two aptitude tests have been applied, one to fourth grade and the other to sixth grade pupils. Designate the tests as IV and VI. Now IV may have $\bar{x} = 40$ and $s = 10$, while VI has $\bar{x} = 60$, $s = 20$. If pupil A scores 60 in IV and his brother B is graded 80 in VI, their standard deviates (or standard measures),

$$A: x' = \frac{60 - 40}{10} = 2, \quad B: x' = \frac{80 - 60}{20} = 1,$$

show their relative standing in the tests.

It is clear that every sample of standard deviates has $\bar{x}' = 0$ and $s' = 1$. Take for illustration the set,

$$X: 6, 6, 2, 8, 6, 8, 6$$

Since the mean is 6, the deviations are

$$x: 0, 0, -4, 2, 0, 2, 0$$

Now, $s = 2$, so the set of standard deviates is

$$x': 0, 0, -2, 1, 0, 1, 0,$$

in which $\bar{x}' = 0$ and $s' = 1$.

If you now write the regression of Y on X in the form (section 6.5),

$$\hat{y} = bx = r \cdot \frac{s_y}{s_x} \cdot x,$$

then decide to use standard measure for each variate, you will have the simple form,

$$\hat{y} = rx',$$

because $s_y = s_x = 1$, and r is therefore the same as b . Thus, in standard measure r is the regression coefficient, the average change in y' per standard deviation change in x' . We have already noticed that r is unaffected by units of measurement. Now it appears that if we use standard measure, also independent of units of measurements, the distinction between the correlation and regression coefficients disappears.

7.5—Sampling variation of the correlation coefficient. Common elements. A convenient way to draw samples from a normal bivariate population is by use of an old device called *common elements* (7). You may go back to the random sampling scheme of section 3.3, or to the samples already drawn from table 3.1. In a new table, such as 7.2, record some convenient number, say three, of the random pig gains. These gains, or

TABLE 7.2
CALCULATION OF THREE PAIRS OF VALUES OF THE VARIATES X_1 AND X_2 HAVING
COMMON ELEMENTS

The elements are pig gains from table 3.1

Pair	Elements	
1	23	← common → { 23
	44	
	43	← different → { 43
	37	
	<u> </u>	<u> </u>
	$X_1 = 147$	$173 = X_2$
2	40	← common → { 40
	16	
	19	← different → { 16
	30	
	<u> </u>	<u> </u>
	$X_1 = 105$	$117 = X_2$
3	23	← common → { 23
	38	
	37	← different → { 38
	30	
	<u> </u>	<u> </u>
	$X_1 = 128$	$170 = X_2$

elements, are written twice in the table. Then continue the drawing, adding, for example, one more randomly drawn gain to the lefthand column, and two more to the right. The sums constitute the paired values of X_1 and X_2 . Three such pairs are computed in the table. It is clear that there is a relation between the two sums in each pair. If the three common elements all happen to be large, then both X_1 and X_2 are likely large, irrespective of the extra elements contained in each. Naturally, owing to the noncommon elements, the relation is not perfect. If you continue this process, drawing a hundred or more pairs, then compute the correlation, you will get a value of r not greatly different from the population value,

$$\rho = 3/\sqrt{(4)(5)} = 0.67$$

The numerator of this fraction is the number of common elements, while the denominator is the geometric mean of the total numbers of elements in the two sums, X_1 and X_2 . A formula may be written for this parameter. If n_{12} represents the number of common elements, with n_1 and n_2 designating the total numbers of elements making up the two sums, then the correlation between these two sums is theoretically,

$$\rho = n_{12}/\sqrt{n_1 n_2}$$

Of course, there will be sampling variation in the values calculated from drawings. You may be lucky enough to get a good verification with only 10 or 20 pairs of sums. With 50 pairs we have usually got a coefficient within a few hundredths of the expected parameter, but once we got 0.28 when the population value was

$$n_{12}/\sqrt{n_1 n_2} = 6/\sqrt{(9)(16)} = 0.5$$

If you put the same numbers of elements into X_1 and X_2 , then $n_1 = n_2$. Denoting this common number of total elements by n , the parameter may be symbolized by

$$\rho = n_{12}/n,$$

the ratio of the number of common elements to the total number in each sum. In this special case, the correlation coefficient is simply the fraction of the elements which are common. Roughly, this is the interpretation of the sister-brother correlation in stature (table 7.1), usually not far from 0.5: an average of some 50% of the genes determining height are common to sister and brother.

Another illustration of this special case arises from the determination of some physical or chemical constant by two alternative methods. As an example, consider the estimation of the potassium content of the expressed sap of corn stems as made by two methods, the colorimetric and the gravimetric (16). Two samples are taken from the same source, one being treated by each of the two techniques. The common elements in the two results must be associated with the actual weights of potassium in the

pair of samples. Extraneous elements may be attributed to the differences in the accuracy with which the potassium is determined by the two procedures, as well as to the random variation in the two samples. The correlation between 24 pairs of determinations was 0.87. On the assumption that each determination is the sum of some constant number of elements chosen at random from a homogeneous population, it may be asserted that 87% of the elements in the two determinations are common. Thus, 87% of the elements are associated with the actual weight of potassium in the sap, the remaining 13% being attributed to errors of measurement and sampling variation.

A less specific illustration is contained in the correlation, $r = 0.70$, between the weights of cob and grain in 250 ears of Country Gentleman sweet corn (11). If you are willing to postulate an equal number of elements determining cob weight and grain weight, then 70% of them may be common. But what are they? Some of them inherited, perhaps, and some environmental.

The reader will have no difficulty in extending this list of illustrations. The disturbing necessity is to keep in mind the limitations. The elements in the two variates must not only be equal in number, but must be drawn from a normally distributed population. Further, the effects of the elements must be describable by a sum. If one element, included by chance, should double the effects of other elements, then the interpretation would be faulty. Despite these limitations, the careful thinker can greatly enhance his ideas of correlation by this concept of common elements.

Unless there is the same unit of measurement in X_1 and X_2 , there is little hope of applying the theory of common elements. For instance, in Dungan's study of hail injury to maize (example 6.4), while both the variates are percentages, yet one is based on areas, the other on weights. It would be hard to specify the nature of the common elements, though undoubtedly there are some.

When you have carried through one or more of the calculations of r with common elements, you are well aware of the sampling variation of this statistic. However, it would be too tedious a task to compute enough coefficients to gain even a rough picture of the distribution curve. This must be left to the mathematicians to be done from theoretical considerations. In figure 7.3 are the curves for samples of 8 drawn from populations with correlations zero and 0.8. Even the former is not quite normal. The reason for the pronounced skewness of the latter is not hard to see. Since the parameter is 0.8, sample values can exceed this by no more than 0.2, but may be in defect of it by as much as 1.8. Wherever there is a limit to the variation of a statistic at one end of the scale, with practically none at the other, the distribution curve is likely to be asymmetrical. Of course, with increasing sample size this skewness tends to disappear. Samples of 300 pairs, drawn from a population with a correlation even so great as 0.8, would have little tendency to range more than 0.05 on either

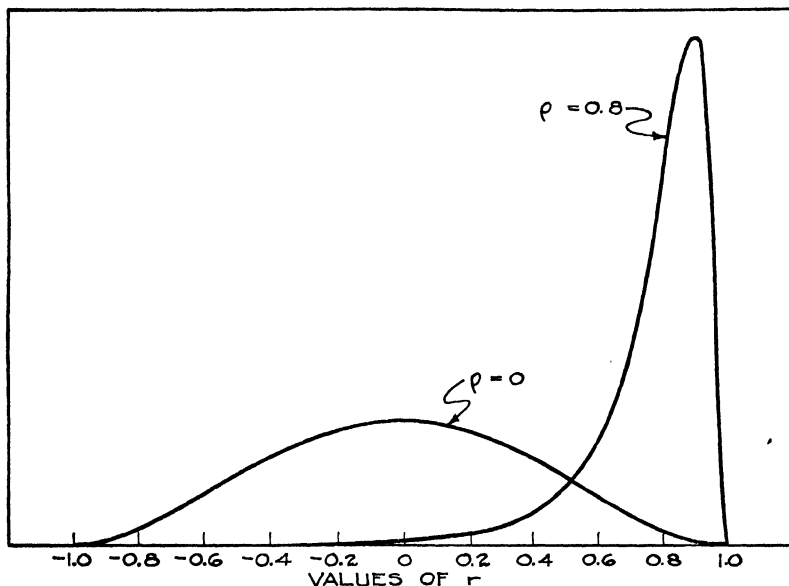


FIG. 7.3—Distribution of correlation coefficients in samples of 8 pairs drawn from two normally distributed bivariate populations having the indicated values of ρ .

side of the parameter. Consequently, the upper limit, unity, would not constitute a restriction, and the distribution would be almost normal.

EXAMPLE 7.9—Suppose all the elements of X_1 are included in X_2 . Then $n_{12} = n_1$ and $\rho = \sqrt{n_1/n_2}$. This is the relation that may exist when the whole, X_2 , is correlated with one of its parts, X_1 . An example is the correlation between the live weight and warm dressed weight of swine (1). Among the 533 animals studied, the mean live weight was 222.6 pounds while the mean dressed weight was 185.1 pounds. If one assumes that each pound of dressed weight represents a common element, then r might be an estimate of $\sqrt{185.1/222.6} = 0.91$. Compare this with the sample $r = 0.986$ and the discussion of section 7.11.

EXAMPLE 7.10—In a tea plantation (5), the production of 16 plots during one 14-week period was correlated with the production of the same plots in the following period of equal length. The correlation coefficient was 0.91. Can you interpret this in terms of common elements?

7.6—Estimating the correlation coefficient: fiducial statements and tests of hypotheses. If random samples are drawn from a normal bivariate population with correlation, ρ , the statistic, r , though not unbiased (8), is an appropriate estimate of ρ . Errors in either X or Y will further bias the coefficient. Unlike regression samples, neither X nor Y may be selected. You see, then, that in samples such as the soybean heights of table 6.3, the correlation coefficient provides only a doubtful estimate of ρ because X is selected. This is a marked distinction between samplings for estimating β and ρ .

Fisher showed also that the null hypothesis, $\rho = 0$, may be tested by use of t whose sample value is

$$t = r\sqrt{(n-2)/(1-r^2)}, d.f. = n-2$$

For illustration, assume that the $r = 0.597$ in C of figure 7.2 is randomly drawn under the conditions specified above. What is the probability of a larger r in sampling from $\rho = 0$? We have

$$t = 0.597\sqrt{(9-2)/(1-0.597^2)} = 1.969, d.f. = 7$$

Interpolation in the t -table elicits the approximate probability, 0.09, of a larger value of t . Hence, a greater value of r could happen about 9 times per 100 in sampling from $\rho = 0$, and the null hypothesis would not likely be rejected. This throws light on the difficulty (section 7.2) of graphical evaluation of small correlations—they may be no more than accidents of sampling.

The test of this null hypothesis, $\rho = 0$, is made at sight in table 7.3.

TABLE 7.3
CORRELATION COEFFICIENTS AT THE 5% AND 1% LEVELS OF SIGNIFICANCE

Degrees of Freedom	5%	1%	Degrees of Freedom	5%	1%
1	.997	1.000	24	.388	.496
2	.950	.990	25	.381	.487
3	.878	.959	26	.374	.478
4	.811	.917	27	.367	.470
5	.754	.874	28	.361	.463
6	.707	.834	29	.355	.456
7	.666	.798	30	.349	.449
8	.632	.765	35	.325	.418
9	.602	.735	40	.304	.393
10	.576	.708	45	.288	.372
11	.553	.684	50	.273	.354
12	.532	.661	60	.250	.325
13	.514	.641	70	.232	.302
14	.497	.623	80	.217	.283
15	.482	.606	90	.205	.267
16	.468	.590	100	.195	.254
17	.456	.575	125	.174	.228
18	.444	.561	150	.159	.208
19	.433	.549	200	.138	.181
20	.423	.537	300	.113	.148
21	.413	.526	400	.098	.128
22	.404	.515	500	.088	.115
23	.396	.505	1,000	.062	.081

Portions of this table were taken from Table VA in "Statistical Methods for Research Workers" by permission of Professor R. A. Fisher and his publishers, Oliver and Boyd.

Simply look along the row for $d.f. = 7$ and observe the position of the sample r relative to the tabular values. Our $r = 0.597$ is considerably less than the 5% level, 0.666, leading to the same conclusion as before. The test is made without considering the sign of r . Among the following

correlations, observe particularly how conclusions are affected by both sample size and the size of r :

Number of Pairs	Degrees of Freedom	r	Conclusion About Hypothesis, $\rho = 0$
20	18	0.60	Reject at 1% level
100	98	0.21	Reject at 5% level
10	8	0.60	Not rejected
15	13	-0.50	Not rejected
500	498	-0.15	Reject at 1% level

Those who wish a sketch of the main features of statistics at first reading may well omit the remainder of this chapter, together with all of the next. Go to chapter 9 if you wish to learn more of enumeration statistics, or to chapter 10 for the elements of analysis of variance.

In figure 7.3 was shown the asymmetry of distributions of small sample r 's from populations with large values of ρ . Only under the hypothesis, $\rho = 0$, can the t -distribution be used precisely for testing significance. It is unsuited for testing other null hypotheses, such as $\rho = 0.5$ for example, or $\rho_1 - \rho_2 = 0$. Equally, t cannot be used for making fiducial statements about small sample correlations. A convenient and sufficiently accurate solution of these problems was provided by Fisher (9) who devised a transformation from r to a quantity, z , distributed almost normally with variance,

$$s_z^2 = \frac{1}{n-3},$$

"practically independent of the value of the correlation in the population from which the sample is drawn." The relation of z to r is given by

$$z = \frac{1}{2}[\log_e(1+r) - \log_e(1-r)],$$

but figure 7.4 enables us to change from one to the other with sufficient accuracy for ordinary sampling. Following are some examples of the use of z .

1. *It is required to set fiducial limits to the value of ρ in the population from which a sample r has been drawn.* As an example, consider $r = -0.889$, based on 9 pairs of observations, figure 7.2F. From figure 7.4, $z = 1.417$ corresponds to $r = 0.889$. Since $n = 9$, $s_z = \sqrt{1/(9-3)} = 0.408$. Now, z is distributed almost normally, independent of sample size, so $t_{.01} = 2.576$, corresponding to $d.f. = \infty$. This value may be taken from table 3.8, or approximately from the table of probability in the normal distribution, table 8.6. The 99% fiducial limits, then, are $l_1 = 1.417 - (2.576)(0.408) = 0.366$, and $l_2 = 1.417 + 1.051 = 2.468$. Returning to the figure, and restoring the sign, the corresponding limits to ρ are

$$l_1 = -0.350, \quad l_2 = -0.986$$

Emphasis falls on two facts: (i) in small samples the estimate, r , is not very reliable; and (ii) the limits are not equally spaced on either side of r , a consequence of its skewed distribution.

2. *Occasionally there is reason to test the hypothesis that ρ has some particular value, other than zero, in the sampled population* ($\rho = 0$, you recall, is tested by use of table 7.3). An example was given in section 7.5, where $r = 0.28$ was observed in a sample of 50 pairs from $\rho = 0.5$. What is the probability of a larger deviation? For $r = 0.28$ take from the figure $z = 0.288$, and for $\rho = .5$, $z = 0.549$. The difference, $0.549 - 0.288 = 0.261$, has the standard error, $1/\sqrt{n-3} = 1/\sqrt{47} = 0.1459$. Hence, $t = 0.261/0.1459 = 1.80$, with $d.f. = \infty$, does not reach the 5% level: the sample is not as unusual as 1 in 20.

3. *To test the hypothesis that two sample values of r are drawn at random from the same population*, convert each to z , then test the significance of the difference between the two z 's. For two lots of pigs the correlations between gain in weight and amount of feed eaten are recorded in table 7.4.

TABLE 7.4
TEST OF SIGNIFICANCE OF THE DIFFERENCE BETWEEN TWO CORRELATIONS OF GAIN
WITH FEED EATEN AMONG SWINE

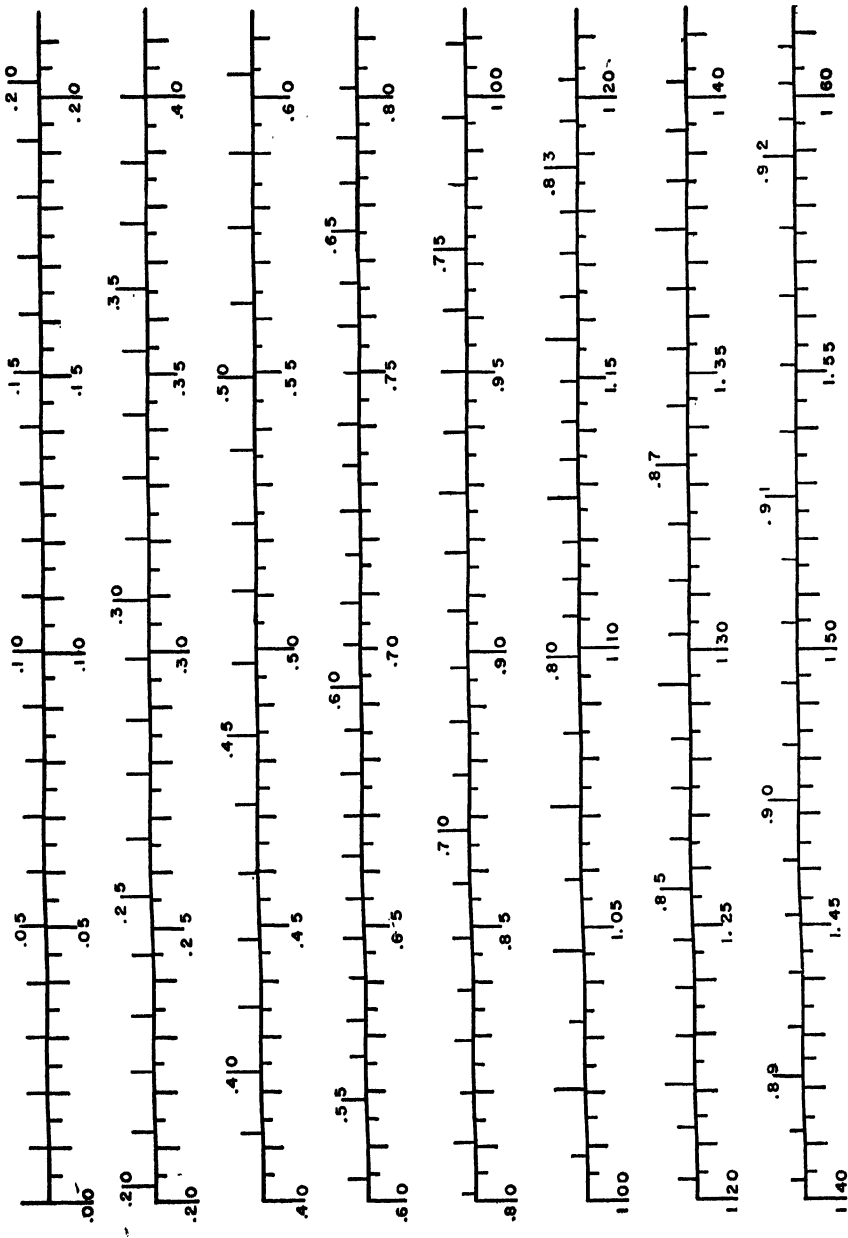
Lot	Pigs in Lot	r	z	$1/(n-3)$
1	5	0.870	1.333	0.500
2	12	0.560	0.633	0.111
Difference = 0.700			Sum = 0.611	
$s_d = \sqrt{0.611} = 0.782$. $t = 0.700/0.782 = 0.895$. $D.f. = \infty$. $P = 0.37$				

The difference between the z -values, 0.700, has the variance,

$$\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3} = \frac{1}{2} + \frac{1}{9} = 0.611$$

The test is completed in the usual manner, calculating t as the ratio of the difference of the z 's to the standard error of this difference. With $P = 0.37$ there is clearly no reason to reject the hypothesis that the z 's are from the same population, and hence that the r 's are from a common population correlation.

4. *To test the hypothesis that several r 's are from the same ρ , and to combine them into an estimate of ρ* . There is often occasion to think that several sample correlations are drawn from a common ρ . If this null hypothesis is not rejected, then it is appropriate to combine the r 's into an estimate of ρ more reliable than that afforded by any of the separate r 's. Lush (15) was interested in an average of the correlations between initial weight and gain in 6 lots of steers. The computations are shown in table 7.5.



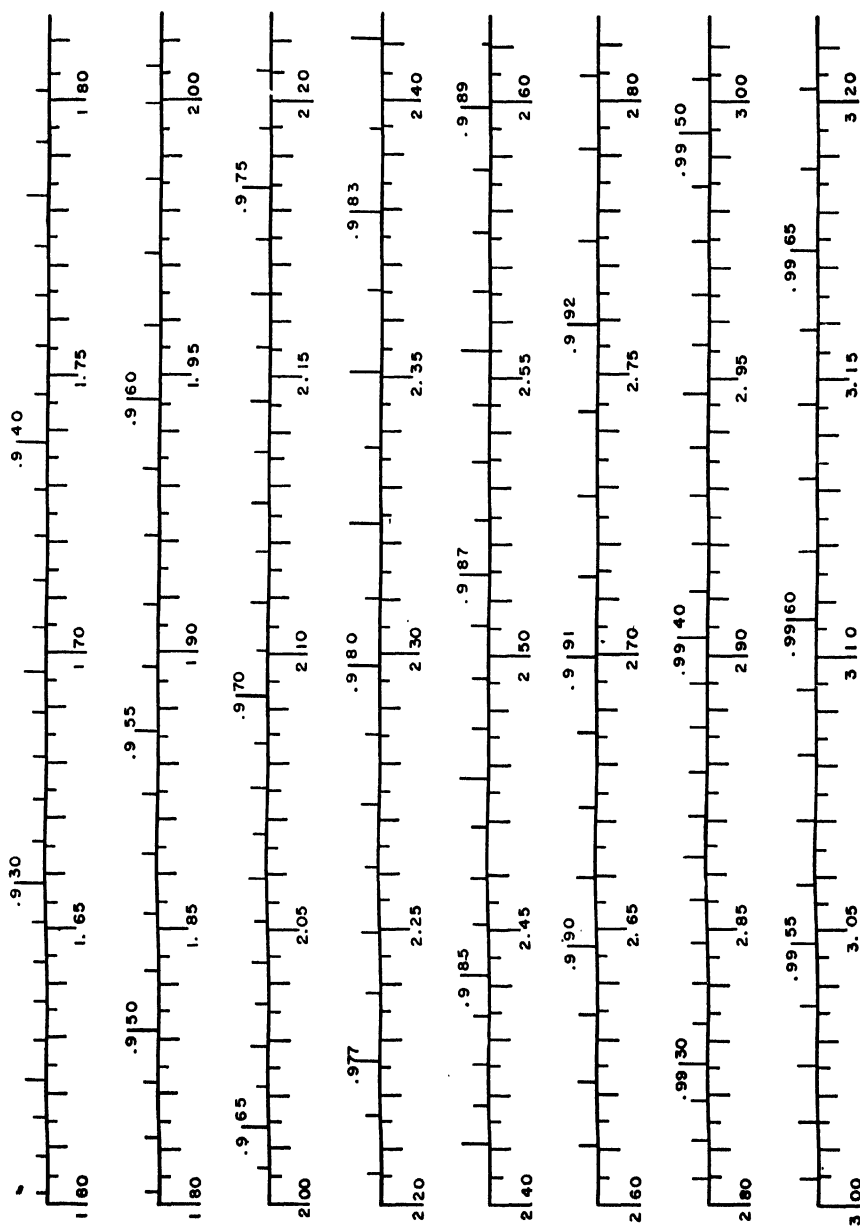


TABLE 7.5

TEST OF HYPOTHESIS OF COMMON ρ AND ESTIMATION OF ρ . CORRELATIONS BETWEEN INITIAL WEIGHT AND GAIN OF STEERS

Samples	No. = n	D.f. = $n - 3$	r	z	Weighted z = $(n - 3)z$	Weighted Square = $(n - 3)z^2$	Cor- rected z
1927 Herefords	4	1	0.929	1.651	1.651	2.726	1.589
1927 Brahmans	13	10	0.570	0.648	6.480	4.199	0.633
1927 Backcrosses	9	6	0.455	0.491	2.946	1.446	0.468
1928 Herefords	6	3	-0.092	-0.092	-0.276	0.025	-0.055
1928 Brahmans	11	8	0.123	0.124	0.992	0.123	0.106
1928 Backcrosses	14	11	0.323	0.335	3.685	1.234	0.321
57		39			15.478	9.753	14.941
				Average $z = 0.397$		6.145	$z = 0.383$
Average $r = 0.377$					$\chi^2 = 3.608$		$r = 0.365$

Each z is weighted (multiplied) by the reciprocal of its variance, so that small samples have little weight. The sum of the weighted z 's, 15.478, is divided by the sum of the weights, 39, to get the average $z = 0.397$. The next column contains the calculations that lead to the test of the hypothesis that the six sample correlations are drawn from a common population correlation. From the sum of the quantities, $(n - 3)z^2$, obtained by multiplying each z by the corresponding $(n - 3)z$, is subtracted a correction,

$$(S(n - 3)z)(\text{average } z) = (15.478)(0.397) = 6.145,$$

the result being a value of chi-square. This value, 3.608, is different from those tabulated in chapter 1 in that it is based on 5 degrees of freedom (less by one than the number of lots) instead of 1. If you look forward to table 9.1 you will find this new distribution in the fifth line, with our present value lying between 3.000 ($P = 0.70$) and 4.351 ($P = 0.50$). Certainly there is no reason to suspect that the 6 correlations may not have been drawn from the same population. This being true, it is appropriate to ask their average value. This is got by taking from figure 7.4 the correlation 0.377 corresponding to the average $z = 0.397$. Don't fail to note the great variation in these small sample correlations.

Fisher pointed out that there is a small bias in z , each being too large by

$$\frac{\rho}{2(n - 1)}$$

You see that the bias usually may be neglected in all the foregoing calculations. It might be serious, however, if large numbers of correlations were averaged, because the bias accumulates, one bit being added with every z . It would be easy to subtract the bias from each z were the population ρ not involved. If there is need to increase accuracy in the calcula-

tion of table 7.5, the average $r = 0.377$, may be substituted for ρ ; then the approximate bias for each z may be deducted, and the calculation of the average z repeated. Since this will decrease the estimated r , it is well to take ρ slightly less than the first average. For instance, it may be guessed that $\rho = 0.37$, then the correction in the first z is $0.37/2(4 - 1) = 0.062$, and corrected z is $1.651 - 0.062 = 1.589$. The other corrected z 's are in the last column of the table. The sum of the products,

$$S(n - 3)(\text{corrected } z) = 14.941,$$

appearing in the calculating machine with no intermediate recording, is divided by 39 to get the corrected mean value of z , 0.383. The corresponding correlation is 0.365. Before straining to reach precision remember two facts: the true ρ is known only approximately, and the values of r and z are read from a graph inaccurate in the third decimal place. For tables of the distribution of r , see reference (4).

EXAMPLE 7.11—To get an idea of how the selection of pairs affects correlation, try picking the 5 lowest values of test II (example 7.5) together with the 6 highest. The correlation between these 11 scores and the corresponding scores on test I turns out to be 0.89, indicating the unreality that can be introduced into r by non-random choice.

EXAMPLE 7.12—Set 95% fiducial limits to the correlation, 0.986, $n = 533$, between live and dressed weights of swine. Ans. 0.983 — 0.988.

What would have been the fiducial limits if the number of swine had been 25? Ans. 0.968 — 0.994.

EXAMPLE 7.13—The correlation, 0.731, between wing and tongue length in bees (example 7.6) was one among 4 similar correlations calculated by Grout, the other three being 0.354, 0.690, and 0.740, each based on a sample of 44. Test the hypothesis that these are samples from a common ρ . Ans. $\chi^2 = 9.164$, $d.f. = 3$, $P = 0.03$. In only about 3 trials per 100 would you expect such disagreement among 4 correlations drawn from a common population. One would like to know more about the discordant correlation, 0.354, before drawing conclusions.

EXAMPLE 7.14—Estimate ρ in the population from which the 3 bee correlations, 0.731, 0.690, and 0.740, were drawn. Ans. 0.721.

EXAMPLE 7.15—Set 99% fiducial limits on the foregoing bee correlation. Note: $r = 0.721$ is based on $3 \times 41 = 123$ $d.f.$ It is therefore equivalent to a single r from a sample of $123 + 3 = 126$ bees. The fiducial limits: 0.590 — 0.815.

EXAMPLE 7.16—Although there is evidence against the hypothesis that the 4 bee correlations were drawn from a common ρ , it would not be appropriate ordinarily to ask if the small one differs significantly from the average of the 3 larger. For sake of illustration, however, let us test the hypothesis that $r = 0.354$, which might have been derived under one set of circumstances, is randomly drawn from the same ρ as the average of the other 3, which could have come from another hive, say. You have found that these 3 larger correlations average 0.721. Show that this differs significantly from the small correlation. Ans. $\chi^2 = 8.97$, $d.f. = 1$, $P < 0.01$.

7.7—Correlation and regression. Every pair of columns in the apple data of table 6.5 will yield a correlation coefficient, some of them of notable interest. Consider the series of equal correlations,

$$r_{XY} = r_{ZY} = r_{Y\hat{Y}} = r_{Y\hat{Y}} = r_{Y\hat{Y}}$$

Each of the corresponding pairs of columns consists of coded values of another pair so that these correlations are all the same (section 7.3).

As examples: (i) The deviations, x and y , are coded values of X and Y got by subtracting $\bar{x}$ from each X and $\bar{y}$ from each Y ; thus, $r_{xy} = r_{XY}$. (ii) Since $\hat{Y} = a + bX$, then $\hat{Y}$ is a coded X , each X being multiplied by b and the product increased by a ; hence, the correlation between X and $\hat{Y}$ is equal to that between $\hat{Y}$ and Y .

This correlation between Y and $\hat{Y}$ is especially enlightening. Restating it, we may say that the correlation between Y and its regression estimate is identical with that between Y and X . This means that if Y is uniformly near the regression line then $r_{Y\hat{Y}}$ and its equal, r_{XY} , are close to ± 1 . Thus, the value of r_{XY} is, in this specific sense, a measure of success in estimating Y by means of regression.

Another correlation worth noticing is the one that you may compute from the columns, X and $\hat{Y}$. Since $\hat{Y}$ is the coded value of X , this correlation between X and its own coded value is ± 1 with sign the same as that of b . Rounding errors may prevent you from verifying this exactly. The meaning of this correlation is that $\hat{Y}$ changes in steps proportional to changes in X , the successive points, $(X, \hat{Y})$, all lying exactly on the regression line.

The correlations of $d_{y \cdot x}$ with x and y may also be calculated and interpreted. The first, $r_{xd_{y \cdot x}} = 0$, recalls one of the assumptions made in fitting this regression: the standard deviation of $d_{y \cdot x}$ is, in the population, uniform throughout the range, and therefore is uncorrelated with x . The second, $r_{yd_{y \cdot x}} = \sqrt{1 - r_{xy}^2}$, may be visualized in this manner: if the regression is zero, the regression line is horizontal, and $d_{y \cdot x}$ coincides with y ; that is, $r_{yd_{y \cdot x}}$ is unity when $r_{xy} = 0$. At the opposite extreme, as the points cluster more and more closely around the line, r_{xy} approaches 1; and since all the deviations, $d_{y \cdot x}$, approach zero, there is nothing for y to keep step with and $r_{yd_{y \cdot x}}$ approaches zero.

It will be observed that in this section r has been treated merely as a calculated quantity without reference to estimation or probability statements. The formulas giving the relation of correlation to regression would be just as true of the soybean data (table 6.3) as of the apple data despite the fact that, on account of the selection of X , the correlation between time and height may not be an informative estimate of the population parameter. This distinction between r as an estimate of ρ and r as a convenient computational device must be observed carefully. Not only in the present section but also in the three to follow, r is used merely as a fraction related to a group of regression concepts. Whether or not it be considered an estimate of ρ depends on whether the sampling conditions of section 7.6 have been met.

EXAMPLE 7.17—Prove that the correlation of $d_{y \cdot x}$ with x is zero. Hint: using deviations, $d_{y \cdot x} = y - bx$. Then the sum of products, $Sxd_{y \cdot x} = Sxy - bSx^2$. Since $b = Sxy/Sx^2$, this reduces to 0.

EXAMPLE 7.18—Prove that the correlation of $d_{y \cdot x}$ with y is $\sqrt{1 - r_{xy}^2}$.

7.8—Correlation and the partition of Sy^2 . The sum of squares attributable to regression (section 6.12) is readily changed as follows:

$$\frac{(Sxy)^2}{Sx^2} = \frac{(Sxy)^2}{Sx^2 Sy^2} \cdot Sy^2 = r^2 Sy^2$$

That is, r^2 is the fraction of Sy^2 due to regression. The remainder is

$$Sd_{y \cdot x}^2 = Sy^2 - r^2 Sy^2 = (1 - r^2) Sy^2$$

You see how intimate is the relation between r and the partitioning of Sy^2 in regression. r^2 is the fraction of Sy^2 associated with the correlated changes in Y and X while $(1 - r^2)$ is the remaining fraction quite independent of X .

If you can bring together these ideas and consider them in relation to figure 6.3 and table 6.5, it is illuminating to study the manner in which this entire complex is affected by different values of r . Take, for example, the extreme case in which $r = 0$, and therefore $1 - r^2 = 1$. None of the variation in percentage wormy apples would then be accounted for by differences in crop size. The regression line would be horizontal, cutting the vertical scale at the point $\bar{y} = 45\%$. The regression percentages in the table would be all the same and equal to $\bar{y}$. The deviations from regression would be simply deviations from mean value. The computation of $s_{y \cdot x}$ would be identical with that of s_y , except for degrees of freedom.

The opposite extreme is that in which $r = \pm 1$ and $1 - r^2 = 0$. Now all the variation in wormy apples would be attributable to regression, every dot in the figure would lie on the regression line, the regression percentages would be identical with the corresponding observed values, and the deviations from regression together with the standard deviation from regression would all be zero.

7.9—Correlation, common elements, and regression. Common elements help to clarify some of the facts of regression. To understand this one must make a slight alteration in the ideas presented in section 7.5. In table 7.2, think of four elements being drawn, their sum constituting X_1 . Transfer three of them, taken at random, to X_2 ; then add two more. The resulting value of X_2 is controlled, to some extent, by the elements transferred from X_1 . In this sense, X_1 is one of the causes of X_2 . While the computation of r is unaffected by this new attitude, the emphasis has been shifted to regression. It has been shown (7) that, using x_1 and x_2 as deviations from mean, the regression of x_2 on x_1 is

$$\hat{x}_2 = (n_{12}/n_1)x_1$$

the regression coefficient being the ratio of the number of transferred elements to the number in x_1 *. As an example, consider the inheritance

* It is convenient to think of x_1 and x_2 as variates instead of X_1 and X_2 . These new variates are the sums you would have got in table 7.2 if you had subtracted $\bar{x} = 30$ pounds from every pig gain in table 3.1.

TABLE 7.6
THREE SAMPLES OF 10 DIFFERENCES, $X_D = X_1 - X_2$, X_1 IS FROM TABLE 3.1, AND X_2 FROM 3.11 WITH EACH ITEM DECREASED 10 POUNDS

Item Number	Sample 1			Sample 2			Sample 3		
	X_1	X_2	X_D	X_1	X_2	X_D	X_1	X_2	X_D
1	39	13	26	19	22	- 3	32	18	14
2	34	19	15	14	20	- 6	31	13	18
3	22	20	2	57	25	32	28	22	6
4	27	34	4	34	26	8	24	17	7
5	33	25	8	39	22	17	44	11	33
6	42	20	22	34	16	18	53	19	34
7	36	7	29	39	23	16	9	16	- 7
8	24	16	8	13	22	- 9	35	23	12
9	25	24	1	39	22	17	33	23	10
10	29	20	9	23	16	7	31	18	13
Mean	31.1	18.7	12.4	31.1	21.4	9.7	32	18	14
Sum of squares	408.9	268.1	918.4	1,706.9	98.4	1,480.1	1,206	146	1,352
Sum of products	-120.7			162.6			0		
Mean square	45.43	29.79	102.04	189.66	10.93	164.46	134.00	16.22	150.22
Covariance	- 13.41			18.07			0		
Correlation	- 0.365			0.102			0		

of stature. An average of half the genes controlling height are transferred to the son by his father. If this were undisturbed by complicating circumstances, one would expect the regression of son's height on father's to be $\bar{x}_2 = 0.5x_1$, the expected deviations of sons' heights from their mean averaging just half the deviations of fathers' heights from theirs. If you are a student of genetics, you will know that this coefficient is often less.

The equation, $\hat{x}_2 = (n_{12}/n_1)x_1$, emphasizes the fact that the estimated value, $\hat{x}_2$, is not affected by the number of elements in x_2 . This estimated value depends entirely upon two quantities: (i) the fraction of the elements transferred from x_1 , and (ii) the value of the sum, x_1 . Neither the number of elements in x_2 , nor their sum, influences $\hat{x}_2$. The values, x_2 and $\hat{x}_2$ are correlated (section 7.7) only because they contain the n_{12} common elements.

7.10—Correlation and differences. Any set of paired observations might be interesting from either or both of two viewpoints; correlation-regression as in chapters 6 and 7, or differences as in chapters 2, 3, and 4. Actually, the two viewpoints are not without connections, some of which we wish to examine now.

In section 3.9 you learned about differences between randomly paired observations from a normal population with $\sigma_1^2 = 100$. Such differences, centering on the mean, zero, had double the variance of the items; that is, $\sigma_D^2 = 2\sigma_1^2$. Again, in section 4.2, corresponding facts about differences between means were pointed out; the differences clustered around zero with variance, $2\sigma_1^2/n$. Now it is necessary to consider differences of a more general type. In the difference, $X_D = X_1 - X_2$, imagine X_1 drawn from a population with mean, m_1 , and variance, σ_1^2 , while X_2 is from m_2 , σ_2^2 . You could draw samples from this population of differences by taking X_1 from the pig gains of table 3.1 and X_2 from those of table 3.11, deducting for convenience 10 pounds from each of the latter items so as to make $m_2 = 20$ pounds. Such differences would be normally distributed with

$$\begin{aligned}m_D &= m_1 - m_2 = 30 - 20 = 10 \text{ pounds} \\ \sigma_D^2 &= \sigma_1^2 + \sigma_2^2 = 100 + 25 = 125\end{aligned}$$

Three samples of this kind are recorded in table 7.6 (the random numbers are the same as those used in table 3.9). Of course, there is sampling variation in both mean and variance as can be seen by comparison of each with its parameter indicated above. This sampling variation leads to a contrast between the relationship among the sums of squares in the sample and in the population. In the sample, it is easily shown (example 7.24) that

$$Sx_D^2 = Sx_1^2 + Sx_2^2 - 2Sx_1x_2$$

Illustrating with sample 1, $918.4 = 408.9 + 268.1 - 2(-120.7)$, and similarly in the other samples. Sample 3 was chosen so that $Sx_1x_2 = 0$, and therefore $1,352 = 1,206 + 146$.

If we substitute for Sx_1x_2 its equal, $r_{12}\sqrt{(Sx_1^2)(Sx_2^2)}$, then divide both members of the equation by $n - 1$, there results,

$$s_D^2 = s_1^2 + s_2^2 - 2r_{12}s_1s_2$$

These equations hold exactly for every set of differences irrespective of sampling variation. Why doesn't the last term appear in the population formula for σ_D^2 ? For this reason: on account of the random pairing of X_1 and X_2 , the correlation in the population of such pairs is zero. Through vagaries of random drawing, a small correlation usually shows up in samples, but the sign is as likely to be plus as minus and the sum of a lot of such correlations would tend towards zero.

If you compare the r 's of samples 1 and 2 with those of table 7.3, you will see that the calculated values may well be samples from $\rho = 0$. Naturally, if you continue the sampling, you will get $|r| > 0.632$ about once in each 20 trials.

The foregoing discussion applies to differences from uncorrelated populations. What happens if X_1 and X_2 are drawn from a population in which there is a correlation, ρ ? Such a population could result from the common elements of section 7.5, and samples can be drawn from it as illustrated in table 7.2: that population, you may recall, had a correlation, $\rho_{12} = 0.67$. If the difference, $X_D = X_1 - X_2$, is calculated from each pair, then

$$\sigma_D^2 = \sigma_1^2 + \sigma_2^2 - 2\rho_{12}\sigma_1\sigma_2,$$

a formula that follows the pattern of that for sample differences from uncorrelated populations. So far as samples are concerned there is no distinction between the formulas. In samples of differences from both correlated and noncorrelated populations,

$$s_D^2 = s_1^2 + s_2^2 - 2r_{12}s_1s_2$$

Of course, if the sample is from a correlated population you may expect r to be significant, especially if ρ and n are fairly large.

This formula for the variance of sample differences explains the effects of arbitrary pairing, examined in section 4.7. The randomly paired differences in the first two columns of table 3.3 are 1, 22, 23, -1, 20, 33, -41, -20, 20, and 6: their variance is, $s_D^2 = 513.79$. In the first column of table 7.7 this variance is verified from the formula above. You will notice that the correlation is negative; hence, the third term in the formula is added to the other two to make up the total, $s_D^2 = 513.79$. Now consider the arbitrary pairing of method I, table 4.8. This arrangement changes the random correlation of -0.59962 to the high positive value, 0.97664 , with the consequent reduction in variance to 8.01 (middle column, table 7.7). The pairing of method II, on the contrary, accentuates the random negative correlation and so increases the variance to $s_D^2 = 627.35$.

The success of pairing in an experiment depends upon the extent to

TABLE 7.7

ILLUSTRATIONS OF THE EFFECT OF CORRELATION ON THE STANDARD DEVIATION OF DIFFERENCES. DATA FROM TABLES 3.3 AND 4.8

	From the First Two Columns of Table 3.3	Method I, Table 4.8	Method II, Table 4.8
r_{12}	— 0.59962 *	0.97664 *	— 0.95353 *
s_1^2	169.82		
s_2^2	151.57		
$s_1^2 + s_2^2$	321.39	321.39	321.39
$2r_{12}s_1s_2$	—192.40	313.38	—305.96
s_D^2	513.79	8.01	627.35

* This number of decimals is necessary to verify the arithmetic.

which positive correlation can be introduced. If the investigator knows his material well enough to bring together into each pair individuals that will react alike, he will tend to get differences like those of method I with a small resultant variation. But if he should try pairing without adequate knowledge of his material, he might get a negative correlation with the large experimental error of method II. In that case, he would have been better off with the group comparison of chapter 4.

Occasionally it may be useful to know that the correlation in a sample of pairs can be calculated from the foregoing formula by rearranging it thus:

$$r_{12} = \frac{s_1^2 + s_2^2 - s_D^2}{2s_1s_2} = \frac{Sx_1^2 + Sx_2^2 - Sx_D^2}{2\sqrt{(Sx_1^2)(Sx_2^2)}}$$

where $x_D = x_1 - x_2$, the deviation of X_D from its mean. An example may be found in the tobacco virus data of table 2.2 where $Sx_D^2 = 130$ is already known. The sums of squares of the numbers of lesions on the two sets of half leaves are easily computed: $Sx_1^2 = 468$ and $Sx_2^2 = 172$.

Substituting,

$$r_{12} = \frac{468 + 172 - 130}{2\sqrt{(468)(172)}} = 0.90$$

The efficiency of this experiment, discussed in section 3.10, arises from the similarity of the reaction of the leaf halves to the inoculation, a similarity reflected in this high correlation, 0.90.

Although the chief interest here is in differences, it is useful to observe, also, the formula for the variance of the sum, $X_S = X_1 + X_2$:

$$s_S^2 = s_1^2 + s_2^2 + 2r_{12}s_1s_2$$

EXAMPLE 7.19—In table 7.1, subtract each sister height from her brother's, then compute the corrected sum of squares of the differences. From this, together with $Sx_1^2 = 74$ and $Sx_2^2 = 66$, verify $r_{12} = 0.558$ by means of the formula just above.

EXAMPLE 7.20—Notice that, in the last formula given above, if $s_D^2 = s_1^2 + s_2^2$ as in sample 3, table 7.6, then $r_{12} = 0$; if $s_D^2 < s_1^2 + s_2^2$ as in sample 2, then r_{12} is positive.

EXAMPLE 7.21—Haber (11) furnished the following information about 300 ears of sweet corn:

Character	Mean (grams)	Standard Deviation (grams)
Ear weight, E	106.55	27.68
Cob weight, C	15.59	
Grain weight, G	90.96	24.62

Observe that $E = C + G$; also, $r_{eg} = 0.9940$. Calculate $s_e = 4.19$ grams, and $r_{eg} = 0.6906$.

EXAMPLE 7.22—If $r_{12} = 1$, show that $s_D = s_1 - s_2$.

EXAMPLE 7.23—If $r_{12} = -1$, show that $s_D = s_1 + s_2$.

EXAMPLE 7.24—Assuming successively that the correlation between X_1 and X_2 is first zero then r_{12} , derive the corresponding formulas for Sx_D^2 and Sx_s^2 .

7.11—Correlations with sums and ratios. Correlations due to common causes. It is often informative to correlate a variate, X_1 , with the sum of itself and another variate, X_2 ; that is, X_1 with $X_s = X_1 + X_2$. As an example, there is the correlation, 0.968 (cited in section 7.3) between warm dressed weight, X_1 , and live weight, X_s , in swine, the variate X_2 denoting the weight of discarded portions. Again, one may gain information from correlating X_1 with $X_p = X_1/X_2$ or with $X_q = X_2/X_1$ (19). An example of r_{1p} is the correlation between the standard deviation and the coefficient of variation; that is, between s and $s/\bar{x}$. For illustration, 50 pairs of these statistics were taken at random from the samples of pig gains, section 3.3. The mean of the 50 standard deviations was 9.87, that of the means, 29.76. The correlation between s and $\bar{x}$ was 0.140, an estimate of $\rho = 0$, because these two statistics are theoretically uncorrelated in this kind of sampling. But the correlation between s and C was 0.932. A familiar instance of r_{1q} is the correlation of live weight, X_1 , with dressing percent, X_q :

$$X_q = \frac{\text{carcass weight, } X_2}{\text{live weight, } X_1}$$

An even more complicated relation is found in the correlation, -0.68 , between average daily gain of swine and feed required per pound gained (6), where the gain is in the numerator of the first rate and in the denominator of the second.

Having observed some unwarranted interpretations of such correlations, Karl Pearson dubbed them "spurious" (17), and this rather derogatory title has led people to distrust them. Of course, it is the interpretation that may be spurious. The correlations are on the same footing as any

others: they are appropriate estimates of parameters if they satisfy the conditions laid down in section 7.6 (1). As for the interpretation, any explanation of a correlation between X_1 and the sums or ratios involving it would seem to require knowledge of r_{12} as well.

An especially simple case, in which $r_{12} = 0$, is illustrated in table 7.8.

TABLE 7.8
CORRELATIONS OF X_1 WITH $X_S = X_1 + X_2$ AND $X_Q = X_2/X_1$. X_1 AND X_2 COMPRISE SAMPLE 3 IN TABLE 7.6 WHERE $r_{12} = 0$

	X_1	X_2	X_1	X_S	X_1	X_Q
	32	18	32	50	32	0.562
	31	13	31	44	31	0.419
	28	22	28	50	28	0.786
	24	17	24	41	24	0.708
	44	11	44	55	44	0.250
	53	19	53	72	53	0.358
	9	16	9	25	9	1.778
	35	23	35	58	35	0.657
	33	23	33	56	33	0.697
	31	18	31	49	31	0.581
Sum	320	180	320	500	320	6.796
Sum of squares	1,206	146	1,206	1,352	1,206	1.59887
Sum of products		0		1,206		-37.516
Correlation		0		0.9445		-0.8543

The correlation, $r_{1S} = 0.9445$, is like the one between dressed weight (X_1) and live weight (X_S) except that here there is no correlation between the two parts. It would be interesting to know the correlation between dressed weight and the weight of discarded parts.

The part-whole correlation can be computed from the original samples by means of the formula,

$$r_{1S} = \frac{s_1 + r_{12}s_2}{\sqrt{s_1^2 + 2r_{12}s_1s_2 + s_2^2}}$$

If $r_{12} = 0$, as in table 7.8, this formula reduces to

$$r_{1S} = \frac{s_1}{\sqrt{s_1^2 + s_2^2}} = \frac{1}{\sqrt{1 + s_2^2/s_1^2}} = \frac{1}{\sqrt{1 + Sx_2^2/Sx_1^2}}$$

The calculation of the part-whole correlation in the table may now be verified by substitution from the first sample:

$$r_{1S} = \frac{1}{\sqrt{1 + 146/1206}} = 0.9445, \text{ as before}$$

If, in addition to $r_{12} = 0$, $s_2 = s_1$ also, then

$$r_{1S} = 1/\sqrt{2} = 0.707$$

This gives some idea of the size of spurious correlations in special cases and shows that both the standard deviations are involved, as well as the correlation between the added variates.

This part-whole correlation may be considered a special case of common elements. X_1 is common to both X_1 and X_s , while X_2 is the different part. If X_1 and X_2 are drawn from the same normal population, then one may put $n_{12} = 1$, $n_1 = 1$, $n_2 = 2$, and each sample correlation is an estimate of

$$\rho = n_{12}/\sqrt{n_1 n_2} = 1/\sqrt{2} = 0.707$$

as in the special case of r_{1s} above.

The correlation, $r_{1Q} = -0.8543$, is negative partly because X_1 is the denominator of the fraction, X_Q , so that a large X_1 tends to make a small X_Q . This correlation, also, can be got directly from the sample. The formula is an approximate one:

$$r_{1Q} = \frac{r_{12}C_2 - C_1}{\sqrt{C_1^2 + 2r_{12}C_1C_2 + C_2^2}},$$

where C is the coefficient of variation, $s/\bar{x}$, expressed as a fraction (not percentage). In the first pair of columns in table 7.8, $r_{12} = 0$, $C_1^2 = 0.13090$, $C_2^2 = 0.05007$ and $C_1 = 0.3618$. Substituting,

$$r_{1Q} = \frac{-0.3618}{\sqrt{0.05007 + 0.13090}} = -0.8505$$

The small discrepancy is due partly to the approximate nature of the formula. In this ratio correlation, the means as well as the standard deviations of the original sample have come into the picture.

The correlation between two variables may be due, wholly or in part, to their common relation to one or more other factors. The organic correlations already mentioned are examples. A big animal tends to be big all over, so that any two parts are likely correlated because of their participation in the general size. Again, two quantities that change in time may show a high correlation. As an example, there is a correlation of -0.98 between the birth rate in Great Britain, from 1875 to 1920, and the production of pig iron in the United States. The matter was discussed by Yule as a question, "Why do we sometimes get nonsense-correlations between time series (21)?" Among the answers suggested only one is pertinent at this point: pairs of observations ordered in time are not randomly drawn in the usual sense and may not come from any normal bivariate population. Hence, they may not conform to the conditions we have set for estimates and probability statements. For another answer, see "partial correlation," chapter 13.

7.12—Correlation of ranks. Occasionally individuals are not measured, or cannot be, but are only arranged in order according to some criterion. Instead of a measurement, each object has a *rank*—1st, 2nd,

etc. Two such rankings of the same things, or of related things, led Spearman (20) to devise a formula for *rank correlation*,

$$r_s = 1 - \frac{6Sd^2}{n(n^2 - 1)},$$

whose calculation is explained in table 7.9. Like r , the rank correlation

TABLE 7.9
RANKING OF 7 RATS BY 2 OBSERVERS OF THEIR CONDITION AFTER 3 WEEKS
ON A DEFICIENT DIET

Rat Number	Ranking by		Difference, d	d^2
	Observer 1	Observer 2		
1	4	4	0	0
2	1	2	-1	1
3	6	5	1	1
4	5	6	-1	1
5	3	1	2	4
6	2	3	-1	1
7	7	7	0	0
			$Sd = 0$	$Sd^2 = 8$

$$r_s = 1 - \frac{6Sd^2}{n(n^2 - 1)} = 1 - \frac{6 \times 8}{7(49 - 1)} = 0.857$$

can range in samples from -1 (complete discordance) to $+1$ (complete concordance). Calculation of r_s in small samples is much easier than that of the product-moment correlation, r .

The utility of rank correlation seems rather limited. Most judges of pigs or bathing beauties feel safer with a score card on which various numbers of points are assigned to each quality considered. Total scores by two or more judges are ordinarily related by use of r . But if there are two rankings to be correlated, Spearman's formula is appropriate.

Little is known about the theory of rank correlation except in two fields: (A) the population in which $\rho = 0$, the two rankings being independent; and (B) the normal bivariate population in which the two coefficients are closely related. References (12), (13), and (14) are excellent accounts of recent thinking about the problem.

A. *The case of independence.* Under the hypothesis, $\rho = 0$, samples of r_s are distributed in a symmetrical fashion that approaches the normal as n increases. The test of significance then involves no assumption about the form of the population. For samples greater than 8 ranks, the test is the same as that of section 7.6, table 7.3, but if n is 8 or less, use table 7.10. One observes that significant evidence against the hypothesis cannot be found in samples of 4 or less. The correlation of table 7.9 is significant at

TABLE 7.10
SIGNIFICANT LEVELS IN SMALL SAMPLES OF THE RANK CORRELATION COEFFICIENT
UNDER THE HYPOTHESIS, $\rho = 0$

Size of Sample	5% Level	1% Level
4 or less	none	none
5	1	none
6	0.886	1
7	0.750	0.893
8	0.714	0.857
9 or more	use table 7.3	

the 5% level but not at the 1%; that is, there are about 2 chances in 100 of getting a rank correlation greater than 0.857 even if the rats were all in the same condition so that the judgments of the observers had no foundation.

B. *Samples from a normal bivariate population.* In this population, ρ and ρ_s are almost the same, the latter being slightly smaller except for perfect or no correlation, in which cases $\rho = \rho_s$. As for samples of r_s from this population, little is known about them unless $\rho = 0$. Using common elements (section 7.5), I drew 10 samples, each with $n = 20$, from a normal bivariate population with

$$\rho = \frac{n_{12}}{\sqrt{n_1 n_2}} = \frac{10}{\sqrt{(20)(16)}} = 0.559,$$

then calculated r and r_s in each. The results are tabulated in table 7.11,

TABLE 7.11
PRODUCT-MOMENT AND RANK CORRELATIONS IN 10 SAMPLES FROM A NORMAL BIVARIATE
POPULATION IN WHICH $\rho = 0.559$

Sample	r	r_s	Sample	r	r_s
1	0.663	0.505	6	0.712	0.680
2	.708	.738	7	.627	.474
3	.505	.584	8	.444	.352
4	.374	.311	9	.529	.398
5	.536	.562	10	.546	.484
Average				0.564	0.512

the averages being got by use of z (section 7.6). In this small sample of coefficients, r_s averages a little smaller than r and is somewhat more variable. The correlation between the 10 pairs of r and r_s is 0.64. There is no strong evidence here that one would not be safe in converting his measurement data to ranks and computing r_s instead of r . For samples of 40 or less, r_s is easier of calculation than is r . However, with good data

from approximately normal populations I would spend the extra time it takes to get the product-moment correlation whose distribution is exactly known.

But suppose one has a sample from a population whose distribution is unknown or is known to be anormal. Then there is some reason to believe that he might be better off with r_s than with r . My advice would be to seek the aid of a mathematical statistician in the hope that a more appropriate method might be found.

7.13—Summary. The correlation coefficient is a measure of co-variation—the degree to which two variates keep in step as they change. The coefficient varies from -1 to $+1$, the extremes indicating perfect linear relationship and zero, independence. The statistic, r , is an estimate of the parameter, ρ , if the sample is drawn randomly from a normal bivariate population: there must be no selection of either variate. Errors in measuring either X_1 or X_2 decrease the precision of the estimate.

The hypothesis, $\rho = 0$, is tested by

$$t = r\sqrt{(n-2)/(1-r^2)}, df. = n-2$$

The transformation, $z = 1.1513 \log \frac{1+r}{1-r}$, almost normally distributed

with variance, $1/(n-3)$, is used for the following: (i) tests of values of ρ different from zero, (ii) fiducial limits, (iii) test of hypothesis that 2 or more samples of r are from the same normal bivariate population, and (iv) averaging two or more values of r on the assumption that they are samples from a common population.

The fraction, r , is often useful in regression calculations even in circumstances where it may not be an estimate of any parameter.

Correlation may be explained variously: by inheritance, by common elements, by organic relations, by the association of two variates with a correlated third variate such as time (nonsense correlation), and by the inherent relationship of sums or ratios (spurious correlation).

Spearman's rank correlation, given by $r_s = 1 - \frac{6Sd^2}{n^3 - n}$, is available

if only rankings are given. It may also be used if there is evidence that the population sampled is anormal.

Intraclass correlation will be presented in chapter 10 and partial correlation in chapter 13.

REFERENCES

1. G. A. BAKER. *Journal of the American Statistical Association*, 37:537 (1942).
2. A. E. BRANDT. Thesis submitted for the degree Doctor of Philosophy, Iowa State College (1932).

3. A. R. CRATHORNE. Reorganization of Mathematics in Secondary Education. Mathematical Association of America, Inc., page 105 (1923).
4. F. W. DAVID. Tables of the correlation coefficient. Cambridge University Press (1938).
5. T. EDEN. Journal of Agricultural Science, 21:547 (1931).
6. JOHN M. EVVARD, M. G. SNELL, C. C. CULBERTSON, and GEORGE W. SNEDECOR. Proceedings of the American Society of Animal Production, 1927, page 2.
7. CARL H. FISCHER. The Annals of Mathematical Statistics, 4:103 (1933).
8. R. A. FISHER. Biometrika, 10:507 (1915).
9. R. A. FISHER. Metron, 1:3 (1921).
10. R. A. GROUT. Iowa Agricultural Experiment Station Research Bulletin No. 218 (1937).
11. E. S. HABER. Data from the Iowa Agricultural Experiment Station.
12. HAROLD HOTELLING and MARGARET RICHARDS PAEST. The Annals of Mathematical Statistics, 7:29 (1936).
13. M. G. KENDALL, S. F. H. KENDALL and B. BABINGTON SMITH. Biometrika, 30:251 (1938).
14. MAURICE G. KENDALL. The advanced theory of statistics, Vol. 1. Charles Griffin and Company, Limited, London (1943).
15. JAY L. LUSH. Journal of Agricultural Research, 42:853 (1931).
16. V. H. MORRIS and R. W. GERDEL. Plant Physiology, 8:315 (1933).
17. KARL PEARSON. Proceedings of the Royal Society, A, 60:489 (1897).
18. KARL PEARSON and ALICE LEE. Biometrika, 2:357 (1902-1903).
19. LOWELL J. REED. Journal of the Washington Academy of Science, 11:449 (1921).
20. C. SPEARMAN. American Journal of Psychology, 15:88 (1904).
21. G. UDN Y YULE. Journal of the Royal Statistical Society, 89:1 (1926).
22. A list of references is given in the Journal of the American Statistical Association, 31:696 (1936).

Chapter 8

LARGE SAMPLE METHODS

8.1—Introduction. The theory of statistics grew up with the notion of a normal frequency distribution. Most of the more fundamental concepts start with the mathematical statement of that same distribution. Its name has been mentioned more than once in this text, but no very definite description of it has been given. In this chapter will be set out so much of the theory as is necessary for the understanding of elementary statistics in biology. The applications are rare but occasionally important. But before looking at the normal distribution in particular, we shall describe some methods of computation appropriate in large samples.

8.2—Computation of mean and standard deviation of a frequency distribution. In section 3.4 you were promised a short method for computing the statistics of a frequency distribution. That of the 511 means is copied in the first two columns of table 8.1. The actual sample means are grouped into classes with midpoints at 19, 20, . . . pounds, the number of means in each class being recorded in the second column. For convenience, the arbitrary origin, G , is selected as 29 pounds because that is the middle class mark and also because that class carries one of the greater frequencies. The code numbers are simply the deviations of the class marks from G . In the fourth column are recorded the sums of the deviations in the several classes. For example, in the fourth class, class mark 22, the deviation is -7 and there are 7 of them: the sum is therefore $(-7)(7) = -49$. At the foot of the column are set down separately the sums of the negative and positive deviations together with their total, 446.

The sums of the squares of the deviations in the last column are calculated conveniently by multiplying fX by X . Again using class mark 22 as an example, $(-7)(-49) = 343$ is easier than $(-7)^2(7)$. Merely multiply twice by each code number, first f to get fX , then fX to get fX^2 . The calculations at the bottom of the table involve no innovations.

Do you see the identity of the code numbers, X , with those discussed in chapter 5? Each mean weight is coded, implicitly, by first subtracting 29, then rounding. Consider a mean whose value was calculated in section 3.4, such as 33.2 pounds. The code is $33.2 - 29 = 4.2$, rounded to $X = 4$. In the actual process of making the frequency distribution the rounding, accomplished by grouping, was the first step. For example, the mean, 33.2 pounds, fell into the class whose mark was 33 pounds. The subtraction of the class mark completed the coding. It is easy to verify

TABLE 8.1
 FREQUENCY DISTRIBUTION OF 511 MEANS OF SAMPLES OF 10 ITEMS (TABLE 3.4).
 COMPUTATION OF MEAN AND STANDARD DEVIATION
 $G = 29$ pounds

Class-mark Pounds	Frequency f	Code Numbers X	Sum of Code Numbers fX	Sum of Squares of Code Numbers fX^2
19	1	-10	-10	100
20	1	-9	-9	81
21	0	-8	0	0
22	7	-7	-49	343
23	5	-6	-30	180
24	10	-5	-50	250
25	19	-4	-76	304
26	30	-3	-90	270
27	41	-2	-82	164
28	48	-1	-48	48
29	66	0	0	0
30	72	1	72	72
31	56	2	112	224
32	46	3	138	414
33	45	4	180	720
34	22	5	110	550
35	24	6	144	864
36	12	7	84	588
37	5	8	40	320
38	0	9	0	0
39	1	10	10	100
$n = \Sigma f = 511$			-444 890	$\Sigma fX^2 = 5,592$
			446	
$\Sigma fX = 446$				$\Sigma fX^2 = 5,592$
$\Sigma fX/n = 446/511$				$(\Sigma fX)^2/n = (446)^2/511 = 389.27$
$= 0.87$				$Sx^2 = 5,202.73$
$\bar{x} = G + \Sigma fX/n$				$s^2 = Sx^2/(n-1) = 5,202.73/510$
$= 29 + 0.87 = 29.87$ pounds				$= 10.2014$
$s_s = \sqrt{s^2/n} = \sqrt{10.2014/511} = 0.141$ pound				$s = 3.194$ pounds

the code numbers for such means as 31.6, 25.1, and 26.8 pounds; namely, 3, -4, and -2. How about 30.5? You remember the rule—round to the adjacent even number. Hence $30.5 - 29 = 1.5$ rounds to 2. But $31.5 - 29 = 2.5$ also rounds to 2. The result is a slight excess of frequency in the even-numbered classes. In effect, these classes are wider than the others.

Another rule must be brought forward from chapter 5, the one requiring the range of code numbers to be not less than 20. In new terms, there must be at least 20 classes for precise work, preferably more. After the computations are done, you can apply this criterion: the standard deviation should be at least four times as great as the class interval. You will notice that this requirement was not fulfilled in table 8.1, s being

only a little more than three times the class interval of one pound. It would have been better to make the class interval one-half or three-quarters of a pound.

It is well to keep in mind that the only lack of precision in using code numbers derives from the grouping or rounding. This process is usually limited to that type of variate known as *continuous*, an item being likely to fall at any point on the scale. There is no sacrifice of precision in the computations when the variate results from such an enumeration as that of the following example.

EXAMPLE 8.1—Edwards (3) planted 20 soybeans on each of 183 agar plates, then observed the numbers germinating. The frequency distribution follows:

Number germinating on each plate, X	6	7	8	9	10	11	12	13	14	15	16
Number of plates, f	5	9	8	19	26	34	26	22	21	10	3

Compute the mean number of beans germinating on a plate, 11.2, together with the standard error, 0.17 bean. Is any inaccuracy introduced by the use of the arbitrary origin? Notice the use of enumeration data as though they were measurements. This is often convenient. If the distribution is normal, one may safely make all inferences appropriate to normal measurement data. For the test of normality, see section 8.6.

8.3—Computation of mean and standard deviation of a frequency distribution—continued. In the last section the unit of measurement, one pound, was conveniently used as the class interval, but in the distribution of table 8.2 it takes an interval of 10 pounds to provide a satisfactory number of classes. The first class interval extends from 75 pounds up to (but not including) 85 pounds, the midpoint being 80 pounds.

EXAMPLE 8.2—Compute the mean and standard deviation of the live weights in table 8.2, taking $G = 180$ pounds.

EXAMPLE 8.3—The yields in grams of 1,499 rows of wheat are recorded by Wiebe (10). They have been tabulated as follows:

Class mark	Fre- quency	Class mark	Fre- quency	Class mark	Fre- quency
375	3	600	127	825	10
400	13	625	140	850	10
425	41	650	122	875	4
450	99	675	94	900	4
475	97	700	64	925	2
500	118	725	49	950	3
525	138	750	31	975	1
550	146	775	26	1,000	1
575	136	800	20		
				Total	1,499

Compute $\bar{x} = 587.74$ grams, and $s = 100.55$ grams. Are there enough classes in this distribution?

EXAMPLE 8.4—Lindstrom (5) reported the numbers of rows of kernels on 327 ears of the F_2 generation derived from a cross of inbred lines of maize :

Number of rows of kernels	12	14	16	18	20	22
Number of ears	25	75	133	68	21	5

The mean is 16.00 rows, the standard deviation 2.136 rows. Is any inaccuracy introduced by the small number of classes?

TABLE 8.2
FREQUENCY DISTRIBUTION OF LIVE WEIGHTS OF 533 SWINE. COMPUTATION OF MEAN AND STANDARD DEVIATION WITH A CLASS INTERVAL DIFFERENT FROM ONE UNIT
Class interval, $I = 10$ pounds $G = 170$ pounds

Class-mark Pounds	Frequency f	Code Numbers X	Sum of Code Numbers fX	Squares fX^2
80	1	-9	-9	81
90	0	-8	0	0
100	0	-7	0	0
110	7	-6	-42	252
120	18	-5	-90	450
130	21	-4	-84	336
140	22	-3	-66	198
150	44	-2	-88	176
160	67	-1	-67	67
170	76	0	0	0
180	55	1	55	55
190	57	2	114	228
200	47	3	141	423
210	33	4	132	528
220	30	5	150	750
230	23	6	138	828
240	11	7	77	539
250	5	8	40	320
260	5	9	45	405
270	4	10	40	400
280	5	11	55	605
290	2	12	24	288

$n = 533$

$-446 \quad SfX^2 = 6,929$
1,011

$SfX = 565$

$SfX = 565$
 $I(SfX)/n = 10(565/533)$
 $= 11$ pounds
 $\bar{x} = G + I(SfX)/n$
 $= 170 + 11$
 $= 181$ pounds

$s_{\bar{x}} = \sqrt{s^2/n} = \sqrt{1,189.86/533} = 1.49$ pounds

$SfX^2 = 6,929$
 $(SfX)^2/n = (565)^2/533 = 598.92$
For code numbers, $Sx^2 = 6,330.08$
 $s^2 = \frac{I^2(Sx^2)}{(n-1)}$
 $= \frac{(10)^2(6,330.08)}{532}$
 $= 1,189.86$
 $s = 34.5$ pounds

G was selected as the midpoint of the class of greatest frequency despite the fact that this was not the middle class. (The size of the correction for mean, 11 pounds, shows that G might well have been taken as 180 pounds.) The class interval, $I = 10$ pounds, enters the calculations of both mean and variance, since both must be decoded.

You will observe that we still haven't provided enough classes. The standard deviation falls short of being four times the class interval, but the discrepancy is not serious.

8.4—Miscellaneous remarks about frequency distributions. There exists some confusion about the desirable number of classes for a frequency distribution. One thinks of two distinct objectives: (i) the presentation of data in summary form, accompanied often by a graph, and (ii) the calculation of the sample statistics. It is only for the latter that 20 or more classes are required. For summarization it is usually desirable to keep the number of classes less than 10.

It requires care to get the class mark in the center of its interval. Suppose weights are recorded in whole pounds with no fractional values. It is the usual practice to estimate such weights to at least the closest half pound; that is, the record of 25 pounds probably included all weights between $24\frac{1}{2}$ and $25\frac{1}{2}$ pounds. If the class intervals are chosen as 20–29 pounds, 30–39 pounds, etc., the first class really extends from 19.5 to 29.5 pounds. In such a scheme, the class mark is $(19.5 + 29.5)/2 = 24.5$ pounds, not 25. Fix the class limits as precisely as feasible, then compute the midpoint to get the class mark.

It is assumed that the items falling into a class are somewhat evenly distributed throughout the interval. When this expectation is realized the mean of the items will be not greatly different from the class mark. This is especially likely in the high frequency intervals bearing the greatest weights in the computations. Caution should be exercised, then, if there are natural groupings in the interval. An instance was observed where the number of locules (seed compartments) in tomatoes was the variate, its values being confined to whole numbers and halves. Only a few of the fruits had half locules. At first, the class intervals were chosen to extend from 2 up to but not including 3, etc., the class marks falling at $2\frac{1}{2}$, $3\frac{1}{2}$, etc. Actually, the mean numbers of locules in the classes were almost at the lower boundaries, 2, 3, etc. This systematic error led to an overstatement of almost half a locule in the mean.

Those desiring to refine the estimate of standard deviation of a continuous variate may use Sheppard's correction for grouping (9). This consists in deducting one-twelfth of a unit from the coded variance, provided the code numbers differ by unity as in tables 8.1 and 8.2. In table 8.1 the correction would be $10.2014 - 0.0833 = 10.1181$ with $s = 3.181$ pounds. This estimate, however, should not be used in a test of significance.

EXAMPLE 8.5—In table 8.2, apply Sheppard's correction to derive the estimate, $s = 34.37$ pounds. Before correction, the coded variance is 11.8986.

EXAMPLE 8.6—Compute the corrected value of s , 100.55 grams, in example 8.3.

EXAMPLE 8.7—In examples 8.1 and 8.4, whose variables are discrete, is there any error of grouping to be corrected?

EXAMPLE 8.8—Two lots of mice inoculated with different isolations of typhoid (see section 11.10 for reference) survived during the numbers of days indicated by the following frequency distributions. Test the significance of the difference between the means of days to death (chapter 4). The sums of squares are 316.88 and 209.64; the

Days to death	1	2	3	4	5	6	7	8	9	10	11	12
Strain 9D	7	5	2	16	11	7	2	5	2	2	0	0
Strain DSC 1	1	0	0	1	6	8	4	6	3	3	1	3

difference between means, 2.628 days, $t = 5.22$, and $d.f. = 93$.

8.5—The test of symmetry. At various points in this text where you have been asked to judge of the normality of sets of data the only criteria available were symmetry and the proportion of items in the interval $x \pm s$. Adequate tests of normality will be described in this and following sections (4). We shall consider two types of departure from the normal. In one, the distribution of the data is asymmetrical or skewed, the mean and median being different. The other occurs in symmetrical sets, characterized by either an excess or a deficit of items concentrated near the center of the range. For the sake of simplicity in presentation, we shall first illustrate the test for symmetry only, using a small sample.

In section 2.17 the array of weights of 11 men was cited as an example of an asymmetrical set. Let us measure the skewness and test significance. The calculations are set out in table 8.3. The first unfamiliar feature is the column of cubes with their sum, $S_3 = 248,628$. Next appear the two k statistics, k_2 being our old friend variance, and k_3 a corresponding average of the third powers of the deviations from mean. The measure of skewness is $g_1 = 1.96$. You will see immediately that S_3 and therefore g_1 might be positive, negative or zero. If g_1 were zero, symmetry in the sample would be demonstrated. A positive g_1 , as in our illustration, indicates an excess in the number of items smaller than the mean. If g_1 proves to be significantly different from zero, the excess of small values may then be ascribed to asymmetry in the population from which the sample is drawn. The standard error of g_1 turns out to be 0.661, so that $t = 2.96$. The test of significance is carried out with an infinite number of degrees of freedom, the 1% level of t being 2.576.

To assist you in following the computations we have carried an unnecessary number of significant figures in table 8.3. For practice, you might recalculate the statistics with the use of no more than four significant figures in any one number.

As a second illustration, let us test the skewness of the weights of

TABLE 8.3
THE TEST OF SYMMETRY IN A SMALL SAMPLE
Weights (pounds) of 11 men

Weight X	Deviation x	Square x^2	Cube x^3
148	-24	576	-13,824
154	-18	324	-5,832
158	-14	196	-2,744
160	-12	144	-1,728
161	-11	121	-1,331
162	-10	100	-1,000
166	-6	36	-216
170	-2	4	-8
182	10	100	1,000
195	23	529	12,167
236	64	4,096	262,144
$\bar{x} = 172$	$S_1 = 0$	$S_2 = 6,226$	$S_3 = 248,628$
$n = 11$	$k_2 = S_2/(n-1)$ $= 6,226/10$ $= 622.6$	$k_3 = nS_3/(n-1)(n-2)$ $= (11)(248,628)/(10)(9)$ $= 2,734,908/90 = 30,387.9$	
$g_1 = k_3/\sqrt{k_2^3} = k_3/(k_2)(\sqrt{k_2})$ $= 30,387.9/(622.6)(\sqrt{622.6})$ $= 30,387.9/15,535 = 1.96$ $s_{g1} = 0.661,$		$s_{g1}^2 = 6n(n-1)/(n-2)(n+1)(n+3)$ $= (6)(11)(10)/(9)(12)(14)$ $= 660/1,512 = 0.437$ $t = g_1/s_{g1} = 1.96/0.661 = 2.96, df = \infty$	

neutral fats in the blood plasma of normal men, using the data recorded in example 2.43. These data call for the introduction of the machine method of calculation. Accumulate the sum of the 64 weights, $s_1 = 14,361$; the sum of their squares, $s_2 = 4,518,741$; and the sum of cubes, $s_3 = 1,878,426,855$. The last sum is obtained most easily by taking the cubes from one of the many available tables containing the first three powers of the integers (1) (2) (6).

From s_2 you compute in the usual manner the sum of the squares of deviations from mean,

$$S_2 = s_2 - s_1^2/n = 4,518,741 - (14,361)^2/64 = 1,296,267,$$

s_1 being the sum of the X 's. For S_3 , the sum of the cubes of deviations from the mean, the corresponding formula is

$$\begin{aligned} S_3 &= s_3 - 3s_1s_2/n + 2s_1^3/n^2 \\ &= 1,878,426,855 - 3(14,361)(4,518,741)/64 + 2(14,361)^3/(64)^2 \\ &= 282,723,308 \end{aligned}$$

Armed with these results, you complete the computations indicated in table 8.3, ending with $g_1 = 1.5696$ and $t = 5.24$. Certainly, there is no doubt of asymmetry in the population from which this sample was drawn. The positive g_1 indicates a mean greater than the median.

correspond to those of tables 8.1 and 8.2. The numbers in the last four columns are calculated by repeated multiplication by X . Take line 3 as an example: $(4)(-6) = -24$, $(-24)(-6) = 144$, $(144)(-6) = -864$, and $(-864)(-6) = 5,184$. You are already familiar with the remaining calculations save those of S_4 , k_4 , g_2 , and s_{g2} . The formulas are intricate but may be followed with a little care.

You notice that both g_1 and g_2 are small, each being exceeded by its standard error. The negative g_1 indicates a slight asymmetry with an excess of items larger than the mean drawing the peak of the frequency curve toward the right. The negative g_2 shows a moderate kurtosis described by a distribution curve plateau-like near the center with an excess of moderate deviations. Since neither g is significant, however, there is little evidence of departure from normality in the population.

EXAMPLE 8.11—Compute the mean ear diameter from table 8.4, together with its standard error, 44.11 ± 0.146 mm.

EXAMPLE 8.12—In table 3.2, compute $g_1 = -0.0139$ and $g_2 = 0.0460$, showing that the distribution is practically normal.

EXAMPLE 8.13—In table 3.6 is the sampling distribution of 511 standard deviations. Calculate $g_1 = 0.3074$ with standard error, 0.108. This indicates that the anticipated departure from normality is significant. g_1 is positive because the mode is to the left of the mean.

EXAMPLE 8.14—The 511 values of t discussed in section 3.7 were distributed as follows:

Class mark	f	Class mark	f	Class mark	f	Class mark	f
-3.13	3	-1.13	29	0.87	31	2.87	1
-2.88	5	-0.88	35	1.12	23	3.12	1
-2.63	1	-0.63	38	1.37	17	3.37	2
-2.38	3	-0.38	40	1.62	11	3.62	0
-2.13	6	-0.13	52	1.87	8	3.87	0
-1.88	12	0.12	57	2.12	10	4.12	0
-1.63	21	0.37	43	2.37	6	4.37	1
-1.38	16	0.62	37	2.62	2		
						Total 511	

The highly significant value of $g_2 = 0.5340$ shows that the frequencies near the mode and in the tails are greater than in the normal distribution, those in the flanks being less. This was expected. But $g_1 = 0.1356$ is nonsignificant, which is also expected because the theoretical distribution of t is symmetrical.

8.7—The graph of the normal curve. This ideal curve follows a famous mathematical formula (example 8.25), the graph of which is shown in figure 8.1. The abscissa is standard measure, t ; that is, deviation from mean expressed in terms of the population standard deviation as a unit. The ordinates are taken from table 8.5. As an example, at $t = 1$ the ordinate is 0.2420. The same ordinate is plotted at $t = -1$. As another example, at $t = 2.46$ the ordinate is read from the column headed 0.06 in the table, opposite the entry $t = 2.4$ in the left column, the length

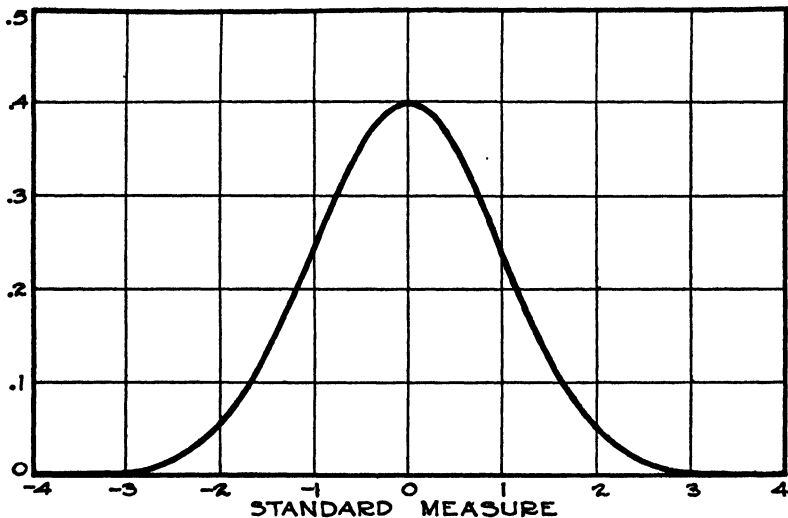


FIG. 8.1—The normal frequency curve. Abscissas are standard measure. Ordinates are taken from table 8.5.

being 0.0194. You will notice that as the curve recedes from the origin it descends faster and faster until it reaches the ordinate at $t = 1$. Beyond that point, the descent is ever slower.

It is easy to identify the ordinates in the table with those in such a distribution as that of ear diameters in table 8.4. You may have already constructed the histogram for the sample. Applying Sheppard's correction to the variance k_2 in table 8.4, you have $s = \sqrt{k_2 - 0.0833} = \sqrt{6.8946} = 2.626$ mm. as the suitable estimate of the value of this parameter. There remains only a constant multiplier to transform the ordinates in table 8.5 to those appropriate for the sample. It is

$$\frac{n}{s} = \frac{327}{2.626} = 124.5$$

Start with the ordinate at the mean, 44.11 mm. Its value is $(124.5)(0.3989) = 49.66$ ears, the second factor being taken from the table at $t = 0$. This ordinate is plotted on the same scale of frequencies used in making your histogram. Next, compute the ordinate at one of the class marks, say 39 mm. The standard measure of this point is $t = (39 - 44.11)/2.626 = -1.95$. The tabular ordinate corresponding is 0.0596; whence the ordinate to be plotted is $(124.5)(0.0596) = 7.42$ ears. As a final illustration, compute the ordinate at one of the class boundaries, say 47.5 mm. You get $t = 1.29$, and thence 21.61 ears as the ordinate to be plotted. Proceeding in this manner, you compute as many ordinates as

TABLE 8.5
ORDINATES OF THE NORMAL CURVE

<i>t</i>	Second decimal place in <i>t</i>									
	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.3989	.3989	.3989	.3988	.3986	.3984	.3982	.3980	.3977	.3973
0.1	.3970	.3965	.3961	.3956	.3951	.3945	.3939	.3932	.3925	.3918
0.2	.3910	.3902	.3894	.3885	.3876	.3867	.3857	.3847	.3836	.3825
0.3	.3814	.3802	.3790	.3778	.3765	.3752	.3739	.3725	.3712	.3697
0.4	.3683	.3668	.3653	.3637	.3621	.3605	.3589	.3572	.3555	.3538
0.5	.3521	.3503	.3485	.3467	.3448	.3429	.3410	.3391	.3372	.3352
0.6	.3332	.3312	.3292	.3271	.3251	.3230	.3209	.3187	.3166	.3144
0.7	.3123	.3101	.3079	.3056	.3034	.3011	.2989	.2966	.2943	.2920
0.8	.2897	.2874	.2850	.2827	.2803	.2780	.2756	.2732	.2709	.2685
0.9	.2661	.2637	.2613	.2589	.2565	.2541	.2516	.2492	.2468	.2444
1.0	.2420	.2396	.2371	.2347	.2323	.2299	.2275	.2251	.2227	.2203
1.1	.2179	.2155	.2131	.2107	.2083	.2059	.2036	.2012	.1989	.1965
1.2	.1942	.1919	.1895	.1872	.1849	.1826	.1804	.1781	.1758	.1736
1.3	.1714	.1691	.1669	.1647	.1626	.1604	.1582	.1561	.1539	.1518
1.4	.1497	.1476	.1456	.1435	.1415	.1394	.1374	.1354	.1334	.1315
1.5	.1295	.1276	.1257	.1238	.1219	.1200	.1182	.1163	.1145	.1127
1.6	.1109	.1092	.1074	.1057	.1040	.1023	.1006	.0989	.0973	.0957
1.7	.0940	.0925	.0909	.0893	.0878	.0863	.0848	.0833	.0818	.0804
1.8	.0790	.0775	.0761	.0748	.0734	.0721	.0707	.0694	.0681	.0669
1.9	.0656	.0644	.0632	.0620	.0608	.0596	.0584	.0573	.0562	.0551
2.0	.0540	.0529	.0519	.0508	.0498	.0488	.0478	.0468	.0459	.0449
2.1	.0440	.0431	.0422	.0413	.0404	.0396	.0387	.0379	.0371	.0363
2.2	.0355	.0347	.0339	.0332	.0325	.0317	.0310	.0303	.0297	.0290
2.3	.0283	.0277	.0270	.0264	.0258	.0252	.0246	.0241	.0235	.0229
2.4	.0224	.0219	.0213	.0208	.0203	.0198	.0194	.0189	.0184	.0180
2.5	.0175	.0171	.0167	.0163	.0158	.0154	.0151	.0147	.0143	.0139
2.6	.0136	.0132	.0129	.0126	.0122	.0119	.0116	.0113	.0110	.0107
2.7	.0104	.0101	.0099	.0096	.0093	.0091	.0088	.0086	.0084	.0081
2.8	.0079	.0077	.0075	.0073	.0071	.0069	.0067	.0065	.0063	.0061
2.9	.0060	.0058	.0056	.0055	.0053	.0051	.0050	.0048	.0047	.0046
<i>t</i>	First decimal place in <i>t</i>									
	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
3	.0044	.0033	.0024	.0017	.0012	.0009	.0006	.0004	.0003	.0002
4	.0001	.0001	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000

may seem necessary for plotting the normal curve. The final graph will have the same general appearance as figure 8.1.

8.8—Table of the normal frequency distribution. Table 8.6 differs from the usual frequency distribution in one major and several minor respects. The chief difference is that the frequencies (percentages or areas) are cumulative. To illustrate, look in the second column, opposite

and $x = 2$. Doubling this percentage, you have 95.44% of the frequency within the interval $x = \pm 2$. The remaining 4.56% of the frequency lies beyond these limits. In general, the numbers in this table correspond to areas in figure 8.1 between ordinates at 0 and x .

Some minor differences between table 8.6 and an ordinary frequency distribution are: (i) Since t is standard measure, its values are measured in terms of the population standard deviation, σ , as a unit. (The Greek lower case sigma, σ , serves to distinguish this parameter from the statistic, s .) Thus, $t = 2$ means $x = 2\sigma$ or $x/\sigma = 2$; (ii) the values of t are not mid-points of class intervals, but are distances to their outer boundaries; (iii) only the positive values of t are listed, corresponding to the right half of figure 8.1; (iv) values of t are given to two decimal places, the second being read at the head of the column. To illustrate this last point, consider the number 4535 in the column headed 0.08 opposite 1.6 in column one. The interpretation is that 45.35% of the frequency lies between $t = 0$ and $t = 1.68$.

Let us determine the value of t such that 5% of the frequency will lie beyond the limits, $-t$ and $+t$. This implies that 2.5% will lie to the right of t , with 47.50% between 0 and t . From the table, the value of t corresponding to 47.50% is 1.96. This, then, is the 5% level in the normal frequency distribution. In working with large samples, the approximate value $x = 2$ is often referred to as the 5% level, though the more precise value, 1.96, is sometimes met. In table 3.8 you will find this figure 1.96 as the 5% value of t for infinite degrees of freedom. That is because t tends toward normal distribution in very large samples.

In large samples where no distinction is made between degrees of freedom and number of observations, it is customary to test the significance of the difference between two means without pooling the sums of squares. The theorem of section 3.9 is used as before, but the variances of the two means are computed separately, then added. The formula is

$$\begin{aligned} \text{either} \quad & s_{\bar{x}}^2 = s_1^2/n_1 + s_2^2/n_2, \\ \text{or} \quad & s_{\bar{x}}^2 = s_{\bar{x}1}^2 + s_{\bar{x}2}^2 \end{aligned}$$

where s_1^2 and $s_{\bar{x}1}^2$ belong to the first sample, s_2^2 and $s_{\bar{x}2}^2$ to the second. As an example, it is reported that for 841 13-year-old boys the mean height was $\bar{x}_1 = 57.3$ inches with $s_1 = 2.76$ inches, while for 784 girls of the same age, $\bar{x}_2 = 58.6$ and $s_2 = 2.44$ inches. The difference of the means is therefore 1.3 inches with standard error,

$$\sqrt{(2.76)^2/841 + (2.44)^2/784} = 0.129 \text{ inch}$$

Since $t = 1.3/0.129 = 10$ is greater than 2, the superior height of the girls at age 13 may be accepted as a population characteristic.

In the borders of table 8.6 containing values of standard measure, $t = x/\sigma$, you will note the range of this variable in a very large sample, from about -4 to 4 . In smaller samples, the average range decreases as

indicated in table 5.5. In all normally distributed populations standard measures are directly comparable. That is, values of x/σ between zero and 1.4, for example, occur with the same probability, 41.92%, no matter what the actual units of measurement may be.

In example 1.25 we had $\chi^2 = 12.2$, a value far beyond the limits of the table. To evaluate the probability, citation was made to this section. Now, it is shown in mathematical statistics that the square root, χ , of the chi-square in chapter 1 is distributed like t in table 8.6. Hence, $\chi = \sqrt{12.2} = 3.49$ may be used as t to enter the present table. The probability read is 0.4998, corresponding to 0.0002 for values greater than $t = 3.49$. For the reason presented in the next paragraph, this probability is doubled, so that we have $P = 0.0004$ that a more divergent sample will occur in random drawing from the hypothetical population. This is the basis of the statement made in example 1.25.

It is now clear that the class intervals for chi-square in table 1.4 may be taken from table 8.6. As an example, let us verify the fact that 10% of the total frequency lies in the interval from $\chi^2 = 0.0158$ to $\chi^2 = 0.0642$. The corresponding values for $\chi = t$ are 0.1257 and 0.2543. Interpolating these values in the table, the probabilities 0.05 and 0.10 are taken out. Hence, the difference, $0.10 - 0.05 = 0.05$, or 5% of the total frequency occurs in this interval. But, since χ^2 is always positive, the corresponding χ distribution is to be taken over only half of the total frequency. Hence 5% of the whole curve constitutes 10% of the right half. Other illustrations of the use of the table will be given in the examples to follow.

EXAMPLE 8.15—Using table 8.6, show that 92.16% of the items in a normally distributed population lie between -1.76σ and 1.76σ .

EXAMPLE 8.16—Show that 65.24% of the items in a normal population lie between $x = -1.1\sigma$ and $x = 0.8\sigma$.

EXAMPLE 8.17—Show that 13.59% of the items are between $x = \sigma$ and $x = 2\sigma$.

EXAMPLE 8.18—Using table 8.6 show that approximately 50% of the items lie in the interval from -0.6745σ to 0.6745σ . Compare this with section 2.16.

EXAMPLE 8.19—Show that the 1% level is at $x/\sigma = 2.575$. You will find a slightly more precise value in table 3.8 at $n = \infty$.

EXAMPLE 8.20—Verify the statement that the interval $\bar{x} \pm 1.645\sigma$ contains 90% of a normal frequency.

EXAMPLE 8.21—In table 8.4, compute the frequency expected in the classes; (i) from 42.5 to 43.5 mm. (Ans. 45.2), (ii) from 43.5 to 44.5 mm. (Ans. 49.2). Hint: Take $\bar{x} = 44.11$ and $s = 2.626$ mm., then for the class boundary 42.5, $t = (42.5 - 44.11)/2.626 = -0.61$.

EXAMPLE 8.22—If you should continue the process suggested in the foregoing example, you would have an observed and expected frequency for every class. Could you test the significance of the departure from normal by the use of chi-square?

EXAMPLE 8.23—In example 1.23 the value of χ^2 was found to be 8.23. What is the probability of a larger value under the hypothesis of independence? Ans. 0.004.

EXAMPLE 8.24—The statement was made that of 1,000 13-year-old boys, 390 had heights within 1.4 inches of the mean height, 57.3 inches. Show that $s = 2.75$ inches on the assumption of normal distribution.

EXAMPLE 8.25—The mean of 1,000 normally distributed plot yields of corn was 20.50 pounds per plot. It was observed that 200 of the plots yielded 17.75 pounds or less. Estimate the standard deviation of the yields. Ans. 3.27 pounds per plot.

EXAMPLE 8.26—The mean and standard error of the dressing percentages in one large group of swine were $81.7 \pm 0.46\%$; in another, $80.0 \pm 0.37\%$. Show that for testing the mean difference $t = 2.88$, a highly significant value.

EXAMPLE 8.27—The equation of the normal curve is

$$y = \frac{n}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}$$

From this, those with some mathematical training can trace the curve, showing that a point of inflexion occurs at $x = \pm\sigma$. Table 8.6 contains values of the integral,

$$\frac{1}{\sqrt{2\pi}} \int_0^t e^{-t^2/2} dt$$

in which $t = x/\sigma$. Can you integrate this function?

8.9—The calculation of the correlation coefficient in a two-way frequency distribution. The double entry table 8.7 contains in compact form the record of two measurements on each of 327 ears of corn (5). The diameters, X , were grouped into millimeter classes as in table 8.4, the frequencies and code numbers appearing in the bottom rows of the table. The weight classes, Y , are 10 grams each with frequency and code at the right. Correlation in these data is evidenced by the tendency of high frequencies to lie along the diagonal of the table, leaving two corners blank—there are no very heavy ears with small diameters. In each diameter array (column) there is notable variation in weight, but it is not nearly so great as the entire range displayed in the lefthand column. The regression of weight on diameter could be represented by a line following the upward trend of the data. This line would be fitted by the method of least squares to the means of the successive diameter arrays.

The calculations follow patterns already familiar. Those leading to Sx^2 are taken from that part of table 8.4 which is the counterpart of table 8.1. The corresponding calculations for weight are carried through with the data in columns f_y and Y . The results are summarized at the bottom of the table.

The device for arriving at SXY , displayed in the last two columns of the table, has not been explained before. Each number in column SXf_x is calculated from the frequencies in that row together with the code numbers in the bottom line. As examples:

- (i) In the 3rd line: $(3)(4) = 12$
- (ii) In the 4th line: $(1)(2) + (1)(4) = 6$
- (iii) In the 5th line: $(1)(1) + (2)(2) + (1)(3) + (4)(6) = 32$
- (iv) In the 7th line: $(1)(-2) + (3)(-1) + (7)(1) + (3)(3) + (3)(4) = 23$

The products in the last column are got from the two columns preceding. The sum at the end of this column is the desired SXY . You remember

TABLE 8.7
COMPUTATION OF CORRELATION COEFFICIENT IN TWO-WAY FREQUENCY TABLE
Frequency of Occurrence of Ears of Maize having each Diameter and Weight

Weight, Grams	Diameter, Millimeters																f_v	Code Y	SY_f	Product $Y(SY_f)$
	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51				
320													1				1	12	4	48
310												1					1	11	3	33
300													3				3	10	12	120
290													1				2	9	6	54
280											1	2	1		4		8	8	32	256
270									1	1	1	1	2	1		1	7	7	23	161
260							1	3	2	7	3	2	3	1			19	6	23	138
250							3	1	2	4	5	6	2	1		1	11	5	24	120
240							1	1	4	3	4	1					23	4	30	120
230						1	2	6	7	3	1	1	1	2			20	3	26	78
220						4	1	7	5	4	3	3	1		1		23	2	5	10
210						3	2	6	6	4	3	1	2				28	1	4	4
200					1	1	2	5	5	11	2	5	1				29	0	7	0
190					2	4	4	6	3	2	2	4	1				32	-1	22	-22
180					1	2	5	5	1	4	2	4	1				28	-2	-14	28
170					1	1	2	5	1	4	2		2				21	-3	-20	60
160					1	1	1	4	1	2	3				1		19	-4	-15	60
150					3	3	2	1		2	1						14	-5	-15	75
140					2	2	1			1							11	-6	-22	132
130					1	1	1										3	-7	-8	56
120					5	2											9	-8	-33	264
110					1					1							3	-9	-4	36
100					1												2	-10	-5	50
90							1		2								4	-11	-8	88
80							1	1									3	-12	-9	108
70																	1	-13	-5	65
60																	1	-14	-4	56
50																	1	-15	-8	120
f_v	1																327		37	2318
Code $\bar{X}$	1	0	4	7	18	26	28	51	47	49	31	33	19	4	8	1				
	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7				

TABLE 8.7—(Continued)

$SXf_x = 37$	$SYf_y = -46$	$SXY = 2,318$
$SX^2f_x = 2,279$	$SY^2f_y = 7,264$	$(SXf_x)(SYf_y)/n = -5.20$
$(SXf_x)^2/n = 4.19$	$(SYf_y)^2/n = 6.47$	$Sxy = 2,323.20$
$Sx^2 = 2,274.81$	$Sy^2 = 7,257.53$	
$r = \frac{Sxy}{\sqrt{(Sx^2)(Sy^2)}} = \frac{2,323.20}{\sqrt{(2,274.81)(7,257.53)}} = 0.5718$		

that no decoding is necessary for r . Incidentally, the sum of the column, SXf_x , checks s_1 in table 8.4.

EXAMPLE 8.28—Using the data in columns f_y and Y , table 8.7, calculate $Sy^2 = 7,257.53$, together with the mean and standard error, 198.6 ± 2.61 .

EXAMPLE 8.29—Calculate the mean, 44.1, and standard deviation, 2.64, in the 42-millimeter array of diameters, table 8.7.

EXAMPLE 8.30—In the 200-gram array of weights, compute $\bar{x} = 198.6$ and $s = 47.18$.

EXAMPLE 8.31—Compute the regression coefficient of weight on diameter, 1.0213, together with the regression equation, $\hat{Y} = 1.0213X + 154.81$.

EXAMPLE 8.32—Calculate the mean weight of the ears in each of the 16 diameter arrays of table 8.7. Present these means graphically as ordinates with the corresponding diameters as abscissas. Plot the graph of the regression equation on the same figure. Do you get a good fit? Is there any evidence of curvilinearity in the regression of means?

EXAMPLE 8.33—Calculate the mean diameter in each of the 28 weight arrays. Plot these means against the weight class marks. Does there seem to be any pronounced curvilinearity in the regression of these mean diameters on the weight? Can you write the regression equation giving estimated diameter for each weight?

EXAMPLE 8.34—If you are interested in mathematics, you will naturally wish to derive the formula for SXY in table 8.7. This table, in fact, introduces many fascinating problems. You will find an excellent discussion, with bibliography, in (8).

8.10—Some theorems about the correlation coefficient in large samples. With increasing sample size the distribution of r tends toward the normal. Furthermore, there ceases to be any notable numerical differentiation between the use of n , $n - 1$, or $n - 2$ in the calculations or tests of significance. The results are marked simplifications in many of the formulas. Some of them will be introduced here for the sake of the accompanying interpretations of r . In reference (7) many others are available to those who have some training in mathematics.

The formula for variance of estimate, section 6.8, is readily transformed (compare section 7.8) to

$$s_{y \cdot x}^2 = (1 - r^2)s_y^2$$

The corresponding formula for the variance of estimated values of Y is

$$s_y^2 = r^2 s_y^2$$

These formulas emphasize the intimate relation between the correlation coefficient and the concepts of regression. They may be written

$$r^2 = s_y^2/s_{y^2} \text{ and } 1 - r^2 = s_{y \cdot x^2}/s_y^2$$

Thus, r^2 measures the fraction of the variance of Y which is due to regression on X , while $1 - r^2$ measures the remaining fraction independent of X . A small value of r^2 goes hand in hand with a small variation in estimated Y , and with a large variation in errors of estimate. Again, if r^2 is large, regression itself accounts for most of the changes in Y , the errors of estimate being small. Still another explanation can be advanced if you recall the fact, clearly evident in the formula of section 7.4, that r is a measure of the correspondence of variations in two sets of deviations from mean. If these variations keep closely in step then the covariation is large, and the variation of X may be said to account for most of the variation of Y . In that event, the residual, unaccounted-for variation, $s_{y \cdot x^2}$ is small relative to s_y^2 .

If desirable, the fraction, $r^2 = s_y^2/s_{y^2}$ may be written as a percentage. For example, if $r = 0.60$, then $r^2 = 0.36$ and it may be said that 36% of the variance of Y is attributable to regression, leaving 64% unaccounted for.

One must discipline his mental processes if he is to think clearly in terms of variance. The effort is worthwhile. Why not use standard deviation instead? One would expect it to be simpler. Try reducing some of the foregoing statements to corresponding ones about standard deviation. For example, since $r = s_y/s_{y^2}$, the correlation coefficient itself may be thought of as the percentage of the standard deviation which is due to regression. Then $r = 0.60$ means that 60% of s_y is accounted for by the regression of Y on X . That is easy. But here is the hitch: $\sqrt{1 - (0.60)^2}$ or 80% is unaccounted for. The two percentages of standard deviation do not add to 100%, whereas those of variance do. It is more convenient, therefore, to talk about variance; but it is of the utmost importance not to think "standard deviation" when you say "variance." Standard deviation is an objective concept, being measured in familiar units like pounds or inches. Variance is a mathematical concept whose mastery requires thought and experience.

REFERENCES

1. PETER BARLOW. Tables. E. and F. N. Spon, Ltd., London (1921).
2. C. B. DAVENPORT and MERLE P. EKAS. Statistical Methods in Biology, Medicine and Psychology. John Wiley and Sons, Inc., New York (1936).
3. T. I. EDWARDS. Plant Physiology, 9:8 (1934).
4. R. A. FISHER. Statistical Methods for Research Workers. Oliver and Boyd, Edinburgh. See section 13 and Appendix D to chapter III.
5. E. W. LINDSTROM. The American Naturalist, 49:311 (1935).

6. **Mathematical Tables.** Chemical Rubber Publishing Co., Cleveland.
7. **HENRY LEWIS RIETZ.** *Mathematical Statistics.* The Open Court Publishing Co., Chicago (1927).
8. **H. L. RIETZ,** Editor-in-Chief. *Handbook of Mathematical Statistics.* Houghton Mifflin Co., New York (1924).
9. **W. F. SHEPPARD.** *Proceedings of the London Mathematical Society*, 29:353 (1898).
10. **GUSTAV A. WIEBE.** *Journal of Agricultural Research*, 50:331 (1935).

Chapter 9

ENUMERATION DATA WITH MORE THAN ONE DEGREE OF FREEDOM

9.1—Introduction. The chief purpose of chapter 1 was the easy introduction of sample-based inferences about populations. Unbiased estimates, confidence statements and tests of null hypotheses were explained with as little diversion as seemed feasible. Those primarily interested in enumeration statistics were referred to this chapter for further exposition.

In the intervening chapters a good deal has been said about the normal distribution: it is the basis for many of the statistical methods devised for measurement data. The methods of the present chapter, including those of chapter 1, are appropriate for samples from two distributions known as the *binomial* and the *Poisson*. In chapter 16, these distributions will be described in somewhat the same fashion as the normal was in chapter 8. We are now ready to extend the methods of chapter 1 in two ways: to experiments involving more than 1 sample and to samples with more than 2 classes of individuals.

9.2—An experiment with more than 1 sample. It is often both convenient and informative to collect data in several small samples rather than in a single large one. An example is furnished by some experiments on chlorophyll inheritance in maize (13) reported in table 9.1. This series consisted of 11 progenies of heterozygous green plants, self fertilized, segregating into dominant green and recessive yellow, presumably in the ratio of 3 to 1.

As an illustration, consider the first progeny with 122 plants, 98 green and 24 yellow. Under the 3:1 hypothesis the numbers expected are three-fourths of 122 = 91.5 green, and one-fourth, 30.5, yellow; then chi-square (review section 1.8 if necessary) is,

$$\chi^2 = \frac{(98 - 91.5)^2}{91.5} + \frac{(24 - 30.5)^2}{30.5} = 1.85,$$

not an excessive value in random sampling from homogeneous material. Omission of the adjustment for continuity will be explained in section 9.4.

The remaining chi-squares are similarly calculated and recorded in the table. Since most of them are small and all nonsignificant, there is no evidence that the 11 progeny samples may not have been drawn from a 3:1 population. In fact, somewhat more variation might have been

TABLE 9.1

NUMBERS OF GREEN AND YELLOW SEEDLINGS IN 11 PROGENIES OF SELF-FERTILIZED, HETEROZYGOUS GREEN PLANTS, WITH NUMBERS EXPECTED UNDER THE 3:1 HYPOTHESIS, AND VALUES OF CHI-SQUARE

Number of Plants	Green	Yellow	Expected Green	Expected Yellow	Chi-square
122	98	24	91.50	30.50	1.85
149	110	39	111.75	37.25	0.11
86	68	18	64.50	21.50	0.76
55	42	13	41.25	13.75	0.05
71	54	17	53.25	17.75	0.04
179	141	38	134.25	44.75	1.36
150	120	30	112.50	37.50	2.00
36	27	9	27.00	9.00	0.00
91	70	21	68.25	22.75	0.17
53	39	14	39.75	13.25	0.05
111	85	26	83.25	27.75	0.15
Sum of 11 chi-squares					6.54
Total 1,103	854	249	827.25	275.25	3.46

expected: you remember that about 1 random sample among 20 has chi-square greater than 3.841.

A valuable property of chi-square is that sums of sample values are, themselves, distributed in the characteristic chi-square fashion. But such sums have distributions different from that which became familiar in chapter 1. The earlier distribution of chi-square was for a single degree of freedom, while the one for the sum of n chi-squares, each with $d.f. = 1$, has n degrees of freedom. The sum, $\chi^2 = 6.54$, in table 9.1 has $d.f. = 11$, and the probability of a larger may be determined by looking along line 11 in table 9.2. It lies about 7/10 of the interval from 0.9 to 0.8: that is, at approximately $P = 0.83$. Clearly, the deviations from 3:1 in these 11 progenies are less than would ordinarily be encountered in sampling from a single binomial population.

The advantage of the sum of chi-squares is that information in the smaller samples is accumulated in the manner of large samples. To illustrate: the sum of 10 chi-squares might be 27.06, 10 times the 10% point for 1 degree of freedom; yet for 10 $d.f.$ this sum is beyond the 1% point. No one of the small samples might have tested significant, but some uniform tendency toward the larger chi-squares makes the sum quite convincing. On the other hand, if the individual chi-squares tend to be small, the accumulated information makes for stronger evidence of homogeneity. You will see this by supposing that the sum of 10 chi-square is 10 times the 50% point for 1 degree of freedom. Try also, the 30% point.

Returning to the corn seedlings, one may well consider pooling the

TABLE 9.2
VALUES OF CHI-SQUARE*

Df.	Probability of a Larger Value of Chi-square													
	0.99	0.98	0.95	0.90	0.80	0.70	0.50	0.30	0.20	0.10	0.05	0.02	0.01	
1	.000	.001	.004	.016	.064	.148	.455	1.074	1.642	2.706	3.841	5.412	6.635	
2	.020	.040	.103	.211	.446	.713	1.386	2.408	3.219	4.605	5.991	7.824	9.210	
3	.115	.185	.352	.584	1.005	1.424	2.366	3.665	4.642	6.251	7.915	9.837	11.341	
4	.297	.429	.711	1.064	1.649	2.195	3.357	4.878	5.989	7.779	9.488	11.668	13.277	
5	.554	.752	1.145	1.610	2.343	3.000	4.351	6.064	7.289	9.236	11.070	13.388	15.086	
6	.872	1.134	1.635	2.204	3.070	3.828	5.348	7.231	8.558	10.645	12.592	15.033	16.812	
7	1.239	1.564	2.167	2.833	3.822	4.671	6.346	8.383	9.803	12.017	14.067	16.622	18.475	
8	1.646	2.032	2.733	3.490	4.594	5.527	7.344	9.524	11.030	13.362	15.507	18.168	20.090	
9	2.088	2.532	3.325	4.168	5.380	6.393	8.343	10.656	12.242	14.684	16.919	19.679	21.666	
10	2.558	3.059	3.940	4.865	6.179	7.267	9.342	11.781	13.442	15.987	18.307	21.161	23.209	
11	3.053	3.609	4.575	5.578	6.989	8.148	10.341	12.899	14.631	17.275	19.675	22.618	24.725	
12	3.571	4.178	5.226	6.304	7.807	9.034	11.340	14.011	15.812	18.549	21.026	24.054	26.217	
13	4.107	4.765	5.892	7.042	8.634	9.926	12.340	15.119	16.985	19.812	22.362	25.472	27.688	
14	4.660	5.368	6.571	7.790	9.467	10.821	13.339	16.222	18.151	21.064	23.685	26.873	29.141	
15	5.229	5.985	7.261	8.547	10.307	11.721	14.339	17.322	19.311	22.307	24.996	28.259	30.578	
16	5.812	6.614	7.962	9.312	11.152	12.624	15.338	18.418	20.465	23.542	26.296	29.633	32.000	
17	6.408	7.255	8.672	10.085	12.002	13.531	16.338	19.511	21.615	24.769	27.587	30.995	33.409	
18	7.015	7.906	9.390	10.865	12.857	14.440	17.338	20.601	22.760	25.989	28.869	32.346	34.805	
19	7.633	8.567	10.117	11.651	13.716	15.352	18.338	21.689	23.900	27.204	30.144	33.687	36.191	
20	8.260	9.237	10.851	12.443	14.578	16.266	19.337	22.775	25.038	28.412	31.410	35.020	37.566	
21	8.897	9.915	11.591	13.240	15.445	17.182	20.337	23.858	26.171	29.615	32.671	36.343	38.932	
22	9.542	10.600	12.338	14.041	16.314	18.101	21.337	24.939	27.301	30.813	33.924	37.659	40.289	
23	10.196	11.293	13.091	14.848	17.187	19.021	22.337	26.018	28.429	32.007	35.172	38.968	41.638	
24	10.856	11.992	13.848	15.659	18.062	19.943	23.337	27.096	29.553	33.196	36.415	40.270	42.980	
25	11.524	12.697	14.611	16.473	18.940	20.867	24.337	28.172	30.675	34.382	37.652	41.566	44.314	
26	12.198	13.409	15.379	17.292	19.820	21.792	25.336	29.246	31.795	35.563	38.885	42.856	45.642	
27	12.879	14.125	16.151	18.114	20.703	22.719	26.336	30.319	32.912	36.741	40.113	44.140	46.963	
28	13.565	14.847	16.928	18.939	21.588	23.647	27.336	31.391	34.027	37.916	41.337	45.419	48.278	
29	14.256	15.574	17.708	19.768	22.475	24.577	28.336	32.461	35.139	39.087	42.557	46.693	49.588	
30	14.953	16.306	18.493	20.599	23.364	25.508	29.336	33.530	36.250	40.256	43.773	47.962	50.892	

* From table III of reference(8) with permission of Professor Fisher and of his publishers, Oliver and Boyd, Edinburgh.

11 progenies into one sample of 1,103 plants because there is so far no evidence against the 3:1 hypothesis. This is done in the last line of the table. The large value, $\chi^2 = 3.46$, $d.f. = 1$, $P = .07$, attracts attention to an excess of green plants in all but 3 of the progenies, an excess that accumulates to $854 - 827.25 = 26.75$ plants. While chi-square is less than the 5% point, it does suggest something to look out for in other experiments: even though no one of the subsample chi-squares were large, yet a persistent bias might lead to a significant value for the merged data. Thus the combined sample might bring to light a slight deviation from hypothetical ratio, undetected in the smaller portions (see example 9.1).

There is a third feature of tables like 9.1 about which evidence is available—the consistency or inconsistency of the progeny ratios. This is indicated by a third chi-square, the difference between the first two:

	Degrees of Freedom	Chi-square
Total	11	6.54
Pooled	1	3.46
Interaction	10	3.08

This *interaction* or *heterogeneity* chi-square (8)(9) is distributed with degrees of freedom equal to the difference between those for the *total* and *pooled* chi-squares. In other experiments this might be informative for this reason; since deviations are squared, and since the progeny chi-squares therefore fail to distinguish between the positive and the negative, the interaction furnishes a test of discrepancies among them.

To illustrate the presence of interaction, the ratios in table 9.3 were selected with alternate positive and negative deviations. The four sample chi-squares are each large though nonsignificant. The total with 4 degrees of freedom is close to the 2% point, indicating over-all deviation from the

TABLE 9.3
RATIOS SELECTED TO ILLUSTRATE INTERACTION IN SAMPLES FROM A 3:1 POPULATION

Size of Subsample	Number Designated		Deviation From 3:1		Chi-square
	A	B	A	B	
116	78	38	-9	9	3.724
92	76	16	7	-7	2.841
152	106	46	-8	8	2.246
124	101	23	8	-8	2.753
484	361	123			11.56
	Total	4	11.56		
	Pooled	1	0.04		
	Interaction	3	11.52, $P = 0.01$ (approximately)		

3:1 ratio. But the pooled data segregate almost 3:1. Comparison of the 2 probabilities suggests heterogeneity among the progeny ratios, further evidence being supplied by $P = 0.01$ for the interaction chi-square. In such an experiment the pooled data have little meaning—apparently none of the progenies represent 3:1 populations. It may be well to emphasize this point: the pooled data with their corresponding chi-square contain evidence about the hypothetical ratio only if the several sample ratios are homogeneous as evidenced by a small heterogeneity chi-square.

EXAMPLE 9.1—Compute the several chi-squares appropriate to the ratios 98:41, 71:31, 127:52, 61:25, 86:36, the theoretical ratio being 3:1. The total chi-square is 6.72, with 5 degrees of freedom. For the pooled data, $\chi^2 = 6.65$, $d.f. = 1$. This example illustrates the situation in which there is a persistent excess of recessive plants, strong enough to boost the chi-square of the pooled data to significance. Consistency in the progeny ratios is shown by the difference, $\chi^2 = 6.72 - 6.65 = 0.07$, $d.f. = 4$.

EXAMPLE 9.2—Here is another group of ratios for practice: 141:38, 83:28, 98:24, 68:18, 67:24, 77:39. Chi-squares are 1.36, 0.00, 1.85, 0.76, 0.09, and 4.60; only the last indicates significant departures from the 3:1 population ratio. The sum, however, $\chi^2 = 8.66$, $d.f. = 6$, is a moderate value ($P = 0.20$). Unless some defect could be discovered in the technique of handling the last group, its large chi-square would be attributed to the vagaries of sampling. In the pooled data variation is not significantly great since $\chi^2 = 0.21$, $d.f. = 1$. Interaction chi-square, $8.66 - 0.21 = 8.45$ with 5 $d.f.$, is not strong evidence against the hypothesis of sampling from a common 3:1 population.

9.3—An experiment with more than 2 classes of individuals. The 9:3:3:1 hypothesis. With liberation from the restriction of a single degree of freedom, all the theoretical genetic ratios may be tested, both in single samples and in combinations. As examples, two groups of progenies are recorded in table 9.4, each being F_2 plants from a cross between japonica colored maize plants and fine-striped. The seedlings were expected to segregate in 9:3:3:1 ratios. Hence, from the 168 plants in the first group, one would expect nine-sixteenths green, three-sixteenths japonica, etc. The formula for chi-square, section 1.8, specifies the sum of the four component parts:

$$\begin{aligned} (117 - 94.5)^2/94.5 &= 5.36 \\ (26 - 31.5)^2/31.5 &= 0.96 \\ (18 - 31.5)^2/31.5 &= 5.79 \\ (7 - 10.5)^2/10.5 &= 1.17 \end{aligned}$$

$$\chi^2 = 13.28$$

This value of χ^2 is based on three degrees of freedom, one less than the number of subsamples contributing. It is larger than the 1% point, 11.34. There is little chance, therefore, that the seedlings in this group are a sample from a population segregating 9:3:3:1. That is, the sample departs significantly from the expected ratios.

In the second group of progenies, $\chi^2 = 9.82$, $d.f. = 3$, is just at the 2%

TABLE 9.4
*F*₂ SEEDLINGS FROM A CROSS OF JAPONICA WITH FINE-STRIPED MAIZE
 Lindstrom's data (13)
 Expected ratios, 9:3:3:1. Expected numbers in parenthesis

Progeny Numbers	Number of Seedlings	Green	Japonica	Fine-Striped	Combination	Chi-square
1, 4, 5	168	117(94.5)	26(31.5)	18(31.5)	7(10.5)	13.28
2, 3	135	82(76.0)	12(25.3)	33(25.3)	8(8.4)	9.82
Total	303	199	38	51	15	

point in table 9.2. The departure from expected is significant but less pronounced than in the first group.

It is easy to see that the total chi-square, $13.28 + 9.82 = 23.10$, $d.f. = 6$, is highly significant. One may be interested in this fact, also: the segregation of the entire 303 seedlings yields $\chi^2 = 12.37$, $d.f. = 3$, a highly significant departure from expected. Finally, the interaction has $\chi^2 = 23.10 - 12.37 = 10.73$, $d.f. = 6 - 3 = 3$, $P = 0.02$. The evidence is the same no matter what particular question is asked; the plants in the sampled populations do not segregate in 9:3:3:1 ratios according to color.

EXAMPLE 9.3—In one of Bateson's experiments (2) on the phenomenon now known as linkage, sweet peas with blue flowers (*B*) were crossed with others having round pollen grains (*I*); also, plants with red flowers (*b*) with those having long pollen grains (*L*). If these characters were assorted independently, the second generation plants from the cross $BI \times bL$ would have segregated into the four classes: BL , Bl , bL , and bl with ratios 9:3:3:1. Actually, the observed numbers were 226:95:97:1. Compute $\chi^2 = 32.36$, $d.f. = 3$. The hypothesis of independent assortment is clearly refuted. Students of genetics will realize that linkage is the cause of the upset theory.

9.4—The adjustment of chi-square for small expected numbers. The reader will perhaps wonder why the adjustment for continuity has not been applied to the foregoing chi-squares. For all values based on a single degree of freedom it is appropriate and should be used if precision is desired in the evaluation of probability, especially in tables having small expected numbers—5 or less. The reasons for the omission of the adjustment in the foregoing sections are four: (i) no great accuracy in P has been required, (ii) the addition theorems do not apply to adjusted values, (iii) the adjustment described in section 1.13 is not appropriate for 2 or more degrees of freedom, and (iv) none of the expected frequencies dealt with have been smaller than 5.

For unbiased evaluation of probability in tables having a single degree of freedom, adjust for continuity in the manner already explained; but if you wish to use an addition theorem, apply it to unadjusted chi-squares.

Yates' adjustment is intended to correct a bias; that is, a tendency toward an estimate different from the true value. In particular samples the adjusted value may not be so accurate as the unadjusted. Those who wish to test this for individual experiments with a single degree of freedom

will find an exact method explained in (21), (8), and (10). Different conclusions may occasionally be drawn if there are few observations in some cells of the table.

9.5—Testing chi-square if degrees of freedom exceed 30. If the degrees of freedom for chi-square exceed 30, the distribution is not far from normal. Fisher has suggested (8) that, as a good approximation, the distribution of table 8.6 be used with

$$t = \sqrt{2\chi^2} - \sqrt{2(df.) - 1}$$

As an example suppose $\chi^2 = 60$ with $df. = 50$. Then

$$t = \sqrt{2(60)} - \sqrt{2(50) - 1} = 1.005$$

The probability of exceeding this value in one-half of the normal curve is $0.5000 - 0.3425 = 0.1575$. For reasons explained in section 8.8, the customary value of P is double this, or 0.315: clearly, $\chi^2 = 60$ is not unusual in random sampling from the chi-square distribution for 50 $df.$

9.6—A test of independence. After a sample has been divided on the basis of one attribute the individuals in each part are often subdivided according to a second. One question that immediately occurs is, "Are the two attributes related or are they independent?" If there are only two classes of each attribute, the 4 subsamples are conveniently presented in a 2×2 or fourfold table like 9.5, *A*. So far as this sample of farms is concerned (review sections 1.2 and 1.3; also example 1.24) the attributes, tenure status and need for labor, are related because

$$100(290)/480 = 60.4\%$$

TABLE 9.5
SUBDIVISION OF THE SOUTH DAKOTA SAMPLE BETWEEN OWNERS AND RENTERS
A. Observed Frequencies

First Attribute: Ability to Utilize Labor to Increase Pigs	Second Attribute: Tenure Status		Total
	Owner	Renter	
Can	290	280	570
Cannot	190	240	430
Total	480	520	1,000

B. Probabilities Calculated From Border Totals

	Owner	Renter
Can	$(0.57)(0.48) = 0.2736$	$(0.57)(0.52) = 0.2964$
Cannot	$(0.43)(0.48) = 0.2064$	$(0.43)(0.52) = 0.2236$

of the owners were able to utilize extra labor, whereas only 53.8% of the renters could. We wish to get evidence on whether the difference between these two percentages may be attributed to sampling variation or whether it reflects a population difference. It is instructive to note that the sample relation between the attributes is equally evident from the percentage of "cans" who are owners

$$(100)(290)/570 = 50.9\%$$

and of "cannots," 44.2%. Any way one looks at the table it is clear that, in the sample, ownership and ability to use extra labor are associated. Apparently these owners were better supplied with the necessary facilities for raising pigs (fencing, houses, feed, etc.), and could increase this enterprise were labor available.

To get evidence about the population, we set up the hypothesis that in it the two attributes are *independent*. The meaning of this term is clarified by reference to the probabilities in part B of table 9.5. Since the probability that a farmer can use extra labor is $570/1,000 = 0.57$, this number is entered in both columns of line 1. In like manner, the probability that he cannot, $430/1,000 = 0.43$, is put in the second line. In both lines of column 1 is entered the probability, $480/1,000 = 0.48$, that a farmer is an owner, while in both lines of column 2 is put down the probability, 0.52, that he is a renter. Now, the hypothesis is that in the population these probabilities are independent. If they are, then according to probability theory the chance that a record be assigned to any cell of the table is the product of the two probabilities in that cell. Thus the probability of being both a "can" and an owner is $(0.57)(0.48) = 0.2736$ provided the two probabilities are independent. Notice the sum,

$$0.2736 + 0.2064 + 0.2964 + 0.2236 = 1,$$

indicating certainty that a record will fall in some one of the four sub-samples.

The product probabilities lead immediately to the set of expected numbers in table 9.6; that is, numbers expected if the attributes are independent. Simply multiply each by the sample size, 1,000. These expected numbers, along with the four deviations, make up a chi-square index of dispersion having the same distribution as the one studied in chapter 1. The calculations are completed in table 9.6. For chi-square there are now four terms in the summation instead of two. Observe carefully that the expected numbers in each row and column add to the border totals and that the deviations add to zero. Since the squares of the deviations are all equal, the formula for chi-square takes the special form,

$$\chi^2 = (X - m)^2 S(1/m),$$

or in words, chi-square is the square of the deviation multiplied by the sum of the reciprocals of the expected numbers.

TABLE 9.6
CALCULATION OF CHI-SQUARE IN A 2×2 TABLE UNDER
THE HYPOTHESIS OF INDEPENDENCE
The South Dakota Farm Sample of 1,000

Ability to Utilize Labor		Tenure Status		Total
		Owner	Renter	
Can	Observed	290	280	570
	Expected	273.6	296.4	
	Deviation	16.4	-16.4	
Cannot	Observed	190	240	430
	Expected	206.4	223.6	
	Deviation	-16.4	16.4	
Total		480	520	1,000

Expected numbers:

$$\begin{aligned}
 (570)(480)/1,000 &= 273.6^* \\
 (570)(520)/1,000 &= 296.4 \\
 (430)(480)/1,000 &= 206.4 \\
 (430)(520)/1,000 &= 223.6 \\
 \hline
 \text{Total} & \quad 1,000.0
 \end{aligned}$$

$$\begin{aligned}
 \chi^2 &= S(X - m)^2/m \\
 &= \frac{(16.4)^2}{273.6} + \frac{(-16.4)^2}{296.4} + \frac{(-16.4)^2}{206.4} + \frac{(16.4)^2}{223.6} \\
 &= 0.983 + 0.907 + 1.303 + 1.203 = 4.40
 \end{aligned}$$

Or,

$$\begin{aligned}
 \chi^2 &= (16.4)^2 \left\{ \frac{1}{273.6} + \frac{1}{296.4} + \frac{1}{206.4} + \frac{1}{223.6} \right\} \\
 &= (268.96) \{0.016346\} = 4.40
 \end{aligned}$$

* This computation is somewhat shorter than that in table 9.5, B.

Comparison of the calculated chi-square, 4.40, $df. = 1$, with table 9.2 shows that only about 4% of random samples from the hypothetical population have larger values. The hypothesis of independence is suspect. It seems reasonable to reject it, concluding that owners in the population are better able than renters to use extra labor.

Rejection of the hypothesis of independence leads to the conclusion that the sample of 1,000 is really drawn from a mixture of two different populations: a population of owners in which 60.4% is an unbiased estimate of the fraction of those who can use help, and a population of renters in which the estimate is 53.8%. Alternatively it may be said that the two populations are these: (i) farmers who can use extra labor, among whom an unbiased estimate of the proportion of owners is 50.9%; and

(ii) farmers who cannot, with the estimate, 44.2%. For each population confidence limits may be set; also, if the necessary information is at hand, all estimates and limits may be expanded to population totals.

In the formula of table 9.6 the adjustment of chi-square is easily made and should be used as routine procedure even though it would modify decisions only in occasional samples of small size. The deviation from expected, $X - m$, is decreased in absolute value by 0.5, the formula being,

$$\chi^2 = (|X - m| - 0.5)^2 S(1/m)$$

Applying this to the data of table 9.6, adjusted chi-square is

$$(15.9)^2(0.016346) = 4.13$$

It may be well to emphasize the contrasts between this section and those preceding. Here is set up no hypothesis, such as 3:1, about a numerical relation between the frequencies of the attributes. The only postulate is that the incidence of the two sets of attributes is uncorrelated. In the 2-way table, the border totals are made the basis of probabilities that individuals will fall into the several subclasses. It follows that only 1 of the 4 cell frequencies can be filled in at will—the other 3 are fixed by subtraction—so that chi-square is distributed with a single degree of freedom. Finally, since the border totals are used to determine the expected numbers in the cells of the table, there is no counterpart of the total and pooled chi-squares calculated in earlier sections: only the interaction is now in question. The hypothesis of independence is identical with the hypothesis that there is no interaction in the population.

The test of independence is equally a test of the hypothesis that the two samples—those in the 2 rows or those in the 2 columns—are randomly drawn from a common binomial population. This is often referred to, also, as a test of homogeneity.

EXAMPLE 9.4—In one of the experiments reported by Lindstrom 94 plants segregated in the ratio, 50 green to 44 yellow. Test the hypothesis that this sample and the 1,103 in table 9.1 are drawn from a common binomial population. Do not assume the 3:1 ratio, but make the test of independence or homogeneity. Ans. Adjusted chi-square is 26.22.

EXAMPLE 9.5—In the fall of 1943 approximately one in every 1,000 families of Iowa was asked about the canning of fruits or vegetables during the preceding season. Of the 370 rural families enumerated 357 had done canning, while among the 300 city (2,500 or more inhabitants) families 274 had canned. Test the hypothesis that the probability of canning is the same in urban and rural communities. Ans. $\chi^2 = 7.11$.

EXAMPLE 9.6—Set 0.95 confidence limits for the percentage of eastern South Dakota owners who can utilize extra help. Ans. About 56% and 65%. Compare this interval with that for the entire sample 54–60 (section 1.3), got without the additional information of table 9.5.

EXAMPLE 9.7—On the assumption that 25,000 farmers in eastern South Dakota are owners or part owners, expand the foregoing estimate and limits. Ans. An estimated 15,100 owners can utilize extra help with 95% confidence limits, 14,000 and 16,200.

9.7—An alternative formula for chi-square in a fourfold table. Compared to that of table 9.6, this formula is usually easier of computation, especially if a calculating machine is available and if the short-cut methods of multiplication and division are used. The numbers in the four cells of the table, and in the borders, may be symbolized thus,

$$\begin{array}{ccc} a & b & a + b \\ c & d & c + d \\ a + c & b + d & n \end{array}$$

where $n = a + b + c + d$. The formula is then written,

$$\chi^2 = \frac{(ad - bc)^2 n}{(a + b)(c + d)(a + c)(b + d)}$$

Let us apply this to the data in table 9.7, taken from a study of the migration of young people from rural homes in Iowa.

TABLE 9.7
CHILDREN OVER 16 YEARS OF AGE WHO HAVE LEFT THEIR RURAL HOMES.
FROM WAKELEY (20)

Location	Males	Females	Total	Percent Males
Open country	607	726	1,333	45.54
Village	231	266	497	46.48
Total	838	992	1,830	45.79

$$\chi^2 = \frac{(726 \times 231 - 607 \times 266)^2 (1,830)}{(1,333)(497)(838)(992)}$$

The numbers involved will be cumbersome unless only the first four or five digits are retained throughout the calculations. The decimal is easily fixed by dividing both numerator and denominator by some convenient power of 10: that is, by moving all decimals to the left an equal number of places in numerator and denominator so that only small groups of digits are involved in the calculations. In this example, after having evaluated the binomial,

$$726 \times 231 - 607 \times 266 = 6,244,$$

the decimals may be moved a total of eleven places in each term of the fraction:

$$\chi^2 = \frac{(0.6244)(0.6244)(1.830)}{(0.838)(0.992)(1.333)(4.97)} = \frac{0.7135}{5.507} = 0.130$$

This value of chi-square indicates that, in the population, the ratio of the sexes is probably independent of the location of the home. The distinction between open country and village may therefore be abandoned and the totals of males and females used to arrive at an unbiased estimate

of the percentage of males, 45.8%, with confidence limits 43.5% and 48.1% (see chapter 16 for computation of confidence intervals in large samples).

In the formula above, the chi-square adjustment is made by subtracting half the total sample size from the absolute value of the binomial in the numerator before it is squared:

$$\chi^2 = \frac{(|ad - bc| - n/2)^2 n}{(a + b)(c + d)(a + c)(b + d)}$$

Use of the formula is illustrated by the records (table 9.8) of clinical

TABLE 9.8
NUMBERS OF TAKES AND FAILURES OF INOCULATIONS WITH VIVAX MALARIA IN THE MONTHS JANUARY TO MARCH CONTRASTED WITH APRIL TO DECEMBER

Months	Result of Inoculation		Total
	Take	Failure	
January-March	13	6	19
April-December	142	10	152
Total	155	16	171

attacks experienced by 155 white patients therapeutically inoculated by anophelines infected with the McCoy strain of *Plasmodium vivax*, employing sporozoites not exceeding 50 days of age (3).

$$\chi^2 = \frac{(142 \times 6 - 13 \times 10 - 85.5)^2 (171)}{(19)(152)(155)(16)} = 9.67$$

This would indeed be an unusual sample from a homogeneous population. Apparently, the success of inoculation with Vivax malaria is not independent of season.

If adjustment is not made, $\chi^2 = 12.45$, a value with probability of being exceeded only one-fourth that of the adjusted (section 8.8).

EXAMPLE 9.8—Dr. C. H. Richardson of the Iowa Agricultural Experiment Station has furnished the following numbers of aphids (*Aphis Rumicis* L.) alive and dead after being sprayed with two solutions of sodium oleate:

Concentration of Sodium Oleate (percent)	Alive	Dead	Total	Percent Dead
1.10	3	62	65	95.4
0.65	13	55	68	80.9
Total	16	117	133	

Calculate adjusted chi-square, 5.31. Estimate the 95% confidence intervals, 87% to 99%, for the population percentage dead in the higher concentration; and 72% to 92% in the lower.

EXAMPLE 9.9—Here are data on the germination of samples from two bags of blue grass seed.

Bag Number	Number of Seeds Germinating	Number Not Germinating	Total	Percent Germinating
1	340	60	400	85
2	356	44	400	89
Total	696	104	800	87

Calculate adjusted $\chi^2 = 2.49$. Since the two bags were from the same shipment, the difference between the percentages of germination may be attributed to sampling variation, and 87% considered as an unbiased estimate of the percent germination in the two bags. Set 0.99 confidence limits to the population percentage. Ans. 84%–90%.

EXAMPLE 9.10—In 1933 Davis (6) was trying to determine the causes of unusual development of smut (*Ustilago zae*) in corn plants. Since all the plants were inoculated, the chances of infection seemed equal. Bags were placed over the ears of 174 plants just before silking, thus preventing pollination and inhibiting development. Of these plants, 38 were smutted. Among 262 unbagged plants 31 were found smutted on September 15. Chi-square is 7.13, indicating a significant difference in the percentages of smutted plants, attributable to the effects of bagging.

EXAMPLE 9.11—In examining the effects of different sprays in the control of codling moth injury to apples, Hansberry and Richardson (12) counted the wormy apples on each of 48 trees. Two trees sprayed with lead arsenate yielded:

- A: 2,130 fruits, 1,299 or 61% of which were injured
 B: 2,190 fruits, 1,183 or 54% of which were injured

Chi-square = 21.16 is conclusive evidence that on these two trees the chance of injury was different. Yet they were treated in all respects alike. This is characteristic of spray experiments. For some unknown reasons, injuries under identical experimental treatment differ significantly. Hence, it is impossible to compare sprays on single trees, because a difference in percentages of injured apples might be attributable either to the treatments or to the unknown sources of variation. A statistical determination of the homogeneity or heterogeneity of experimental material under identical conditions is called a *test of technique*. Hansberry and Richardson combined a study of technique with their experiment on sprays.

EXAMPLE 9.12—The Bureau of Entomology and Plant Quarantine, United States Department of Agriculture, has been investigating the effectiveness of various treatments for killing cabbage worm loopers. From some unpublished data supplied by Messrs. D. J. Caffrey and C. E. Smith (5), we have taken the records of the numbers of plants found infested after two treatments.

- Treatment A: 200 plants examined; 24 found infested
 Treatment B: 200 plants examined; 9 found infested

Chi-square = 6.47 shows that these samples were not likely taken from the same population. Evidently the treatments caused differences in the numbers of infested plants, treatment B being more effective in killing the loopers.

9.8—Three sets of attributes—the $2 \times 2 \times 2$ table. Not infrequently observations are made on more than two attributes; for example on sex, on treatment and on species. If there are only 2 classes of each attribute, independence may be tested by a method due to Bartlett (1). An example is found in an experiment by Gowen (11) on the susceptibility of strains of mice to different isolations of typhoid organisms, table 9.9.

TABLE 9.9
NUMBERS OF 2 STRAINS OF MICE DEAD AND ALIVE 21 DAYS AFTER INOCULATION
WITH 2 ISOLATIONS OF TYPHOID

Isolations	Strains of Mice			
	RI		Ba	
	Dead	Alive	Dead	Alive
9D	260	46	322	27
11C	164	145	336	10

Each of the fourfold tables may be examined for independence. In the *Ba* strain, for example,

$$\chi^2 = \frac{(336 \times 27 - 322 \times 10 - 695/2)^2(695)}{(349)(346)(658)(37)} = 7.16$$

Clearly the mice in this strain do not react alike to the two isolations of the typhoid organism. This absence of independence has been referred to as interaction: chi-square tests the hypothesis that there is no interaction between survival and the two isolations of the organisms.

In the theory being followed, the *estimate* of interaction is the ratio of a/b to c/d , or alternatively, the ratio of a/c to b/d : each ratio is ad/bc . Among the *Ba* mice, this interaction is $(322)(10)/(27)(336) = 0.3549$. By comparing the interaction ratio with the expression for chi-square, you will see that, aside from the continuity adjustment, if $\chi^2 = 0$ then $ad/bc = 1$. In this sense, the hypothesis tested by chi-square is that in the population the interaction ratio is unity.

One more observation about the fourfold table: if the 4 deviations, $X - m$, are represented by x , we learned that $|x|$ is the same for the 4 classes, and that $\sum x = 0$. Consequently, it is easy to construct a table of expected numbers in which the interaction ratio is 1:

$$\begin{array}{cc} a - x & b + x \\ c + x & d - x \end{array}$$

In verification, refer to table 9.6 in which $|x| = 16.4$. The expected numbers in the 4 classes may be written,

$$\begin{array}{cc} 290 - 16.4 & 280 + 16.4 \\ 190 + 16.4 & 240 - 16.4 \end{array}$$

Furthermore,

$$\frac{(290 - 16.4)(240 - 16.4)}{(280 + 16.4)(190 + 16.4)} = \frac{(273.6)(223.6)}{(296.4)(206.4)} = 1$$

It follows that x could be computed by solving the equation,

$$\frac{(a-x)(d-x)}{(b+x)(c+x)} = 1,$$

which is equivalent to the equation,

$$(a-x)(d-x) = (b+x)(c+x)$$

If x were calculated in this way, the convenient formula for adjusted chi-square would be,

$$\chi^2 = (|x| - 0.5)^2 \left(\frac{1}{a-x} + \frac{1}{b+x} + \frac{1}{c+x} + \frac{1}{d-x} \right)$$

In the $2 \times 2 \times 2$ table, the interaction in one fourfold part may be different from that in the remainder. The *RI* mice, for example, interact to the isolations in the amount,

$$(260)(145)/(46)(164) = 4.997$$

whereas the interaction was 0.3549 for the *Ba* strain. The ratio of these *first order* interactions is known as the *second order* interaction,

$$4.997/0.3549 = 14.08$$

If the table is written in the symbolical form,

$$\begin{array}{cc|cc} a & b & a' & b' \\ c & d & c' & d' \end{array}$$

the second order interaction is

$$\frac{ad}{bc} \frac{a'd'}{b'c'} = adb'c'/a'd'bc = 14.08$$

In order to test the hypothesis that in the population the second order interaction is 1, the value of the deviation, $|x|$, common to all 8 cells of the table, is determined by solving the equation,

$$(a-x)(d-x)(c'-x)(b'-x) = (c+x)(b+x)(a'+x)(d'+x)$$

The equation is easily remembered if it is noticed that in the left member the letters follow a diagonal downward in the first 2×2 table, then across and upward in the second. Substituting the mouse data,

$$(260-x)(145-x)(336-x)(27-x) = (164+x)(46+x)(322+x)(10+x),$$

which reduces to

$$1,310x^3 - 113,303x^2 + 20,540,068x - 317,722,720 = 0$$

Methods for approximating the real root of this cubic may be found in any college algebra. Essentially, they consist in trying various values of x until two are found between which the left member of the equation changes sign. A first approximation may be got by ignoring the first two terms, solving the equation,

$$20,540,068x - 317,722,720 = 0$$

From this, $x = 15$ roughly. If you try 15, 16, and 17 in the cubic, you will find that the left member is $-12,721,440$ for $x = 16$ and $5,149,899$ for $x = 17$. This leads to the improved approximation,

$$\frac{x - 16}{17 - x} = \frac{12,721,440}{5,149,899}$$

Solving, $x = 16.7$. If a still better approximation is desired it will be found that the left member is $-22,868$ for $x = 16.71$ and $155,635$ for $x = 16.72$: the equation,

$$\frac{x - 16.71}{16.72 - x} = \frac{22,868}{155,635}$$

gives $x = 16.711$, a more accurate approximation than is ordinarily needful. Verify the result by substitution in the original equation. Note: x may be positive or negative in other examples.

The 8 expected numbers, $(260 - x), \dots (10 + x)$, are now calculated, whence adjusted chi-square, with 1 degree of freedom, is

$$\begin{aligned} \chi^2 &= (16.711 - 0.500)^2 \left(\frac{1}{243.289} + \dots + \frac{1}{26.711} \right) \\ &= 45.75 \end{aligned}$$

This is very strong evidence that the interaction between survival and isolation in the *Rf* strain is different from that in the *Ba*. Equally it may be said that the interaction between strain and survival differs with isolation; also, though it has little meaning in this illustration, the isolation-strain interaction is not the same for the living and dead.

If, in other circumstances, the investigator is willing to accept the hypothesis of no second order interaction ($adb'c' = a'd'bc$ in the population) it would be appropriate to combine the data into one (or more) of the 3 summary fourfold tables, then test the corresponding first order interaction (example 9.13).

EXAMPLE 9.13—In studying the isolating mechanisms of two species of *Drosophila*, Mayr, and Dobzhansky (16) confined males of *D. persimilis* for several days with females of the same species, pre-conditioning them, then gave them access to females of both their own species and of *D. pseudoobscura*. They then confined *persimilis* males

with *pseudoobscura* females, counterconditioning them, following this with multiple choice as before. The numbers of females inseminated are recorded.

	Pro-conditioned		Counter-conditioned	
	Inseminated	Not	Inseminated	Not
<i>Persimilis</i>	18	14	41	6
<i>Pseudoobscura</i>	5	32	20	32

The equation, $(x + 18)(x + 32)(x + 20)(x + 6) = (x - 5)(x - 14)(x - 41)(x - 32)$, has the approximate root, 0.45, hence chi-square (you can't adjust it) is small, 0.13.

If the hypothesis of negligible second order interaction be accepted, it becomes interesting to examine the interaction in the condensed table.

Females	Inseminated	Not
<i>Persimilis</i>	59	20
<i>Pseudoobscura</i>	25	64

The adjusted chi-square value, 34.5, leaves no doubt that, under the conditions specified, males inseminate the females of their own species more freely.

EXAMPLE 9.14—Under the hypothesis of independence, the probabilities in the 4 cells of a 2×2 table such as table 9.5B, may be written

$$\begin{array}{cc} p_1 & p_2 \\ p_3 & p_4 \end{array}$$

Show that $p_1 p_4 = p_2 p_3$. Correspondingly, in the $2 \times 2 \times 2$ table,

$$\begin{array}{cc} p_1 & p_2 \\ p_3 & p_4 \end{array} \quad \begin{array}{cc} p_5 & p_6 \\ p_7 & p_8 \end{array}$$

verify the fact that $p_1 p_4 p_7 p_8 = p_2 p_3 p_5 p_6$.

9.9—Test of independence in an $R \times C$ table. If one or both of two attributes are recorded in more than two classes, independence may be tested by chi-square with $(R - 1)(C - 1)$ degrees of freedom. Based on the border totals, 2 probabilities of occurrence may be calculated for every cell of the table. Under the hypothesis of independence, the products are probabilities that an observation shall fall in the several compartments, and the probabilities lead to expected numbers, m . Finally, $\chi^2 = S(X - m)^2/m$. The hypothesis of independence is identical with the hypothesis that there is no interaction in the population.

Strand and Jessen (18) examined the distribution of tenancy on various levels of soil fertility in Audubon County, Iowa. One question was, "Is tenancy dependent on fertility; that is, is there a tendency for some forms of tenancy to be prevalent on a particular fertility level?" A sample is summarized in table 9.10.

Calculation of the expected numbers follows one of the patterns already explained. In soil I, for example, the probability of an observation is $152/517 = 0.294003$: this, multiplied in succession by 125, 214, and 178,

TABLE 9.10
NUMBERS OF FARMS ON 3 SOIL FERTILITY GROUPS IN AUDUBON COUNTY, IOWA,
CLASSIFIED ACCORDING TO TENURE

Soil		Owned	Rented	Mixed	Total
I	$\bar{X}$	36	67	49	152
	m	36.75	62.92	52.33	
	$\bar{X} - m$	-0.75	4.08	-3.33	
II	$\bar{X}$	31	60	49	140
	m	33.85	57.95	48.20	
	$\bar{X} - m$	-2.85	2.05	0.80	
III	$\bar{X}$	58	87	80	225
	m	54.40	93.13	77.47	
	$\bar{X} - m$	3.60	-6.13	2.53	
Total		125	214	178	517

$$\chi^2 = S \frac{(x - m)^2}{m} = \frac{(-0.75)^2}{36.75} + \dots + \frac{(2.53)^2}{77.47} = 1.54, \text{ d.f.} = (R - 1)(C - 1) = 4$$

gives the expected numbers, 36.75, 62.92, and 52.33. Verification of the expected numbers of owned farms could be had by multiplication of $125/517 = 0.241779$ by the totals in the right-hand border, but this is usually not necessary because (i) the sum of the expected numbers in each row and column must equal the border total, and therefore (ii) the sum of the deviations in each row and column must be zero: hence, verification is easy. From the deviations and the expected numbers, chi-square is calculated as indicated. Clearly, there is no reason to reject the hypothesis of independence.

Two observations: (i) If one or more cells of the table contribute unusually large values of $(\bar{X} - m)^2/m$, they may warrant special consideration. (ii) Since there is no evidence of interaction, it would be proper to test any hypothesis the investigator might have about the border totals. For example, if he should know that in the state as a whole tenancy is distributed in the ratios 25:45:30, he could test (as in section 9.3) whether his sample is randomly drawn from the state population or if Audubon County has a different pattern from that in the state.

9.10—A special method of computing chi-square in an $R \times 2$ table.

In scientific writings many tables of enumeration data are presented in some such fashion as the first three columns of table 9.11. These data are selected from an experiment in which uniform doses of *Danysz bacillus* were injected into rats. The sizes of the subsamples were dictated by the numbers of animals available at the several dates of injection.

TABLE 9.11
PERCENTAGE OF RATS LIVING AFTER INJECTION WITH DANYSZ BACILLUS

Number in Group	Number Living X	Probability* of Survival p	Products pX
40	9	0.2250	2.0250
12	2	0.1667	0.3334
22	3	0.1364	0.4092
11	1	0.0909	0.0909
37	2	0.0541	0.1082
20	3	0.1500	0.4500
142	$SX = 20$	$\bar{p} = 20/142 = 0.1408$ $\bar{q} = 1 - \bar{p} = 0.8592$	$S pX = 3.4167$
$\chi^2 = \frac{SpX - \bar{p}SX}{\bar{p}\bar{q}} = \frac{3.4167 - 2.8160}{(0.1408)(0.8592)} = 4.97, d.f. = 5, P = 0.43$			

* If you prefer percentages, see example 9.18.

The probabilities of survival, p , in the third column of the table are emphasized. For such tables a convenient method of computing chi-square has been devised (17). (This is sometimes referred to as the Brandt and Snedecor method). The results, of course, are the same as those obtained by applying the method of the foregoing section. The second and third columns might as well refer to the numbers dead, but computation is easier with the smaller probabilities. In the last column the sum of the products is usually accumulated in the dials of a calculating machine with no listing of the separate items.

The results indicate that the probability of death is independent of the group and may therefore be characteristic of the amount of inoculum injected. The experimental technique was rather difficult to control—the rats were probably variable in their reaction to the injected organism, the organism doubtless changed somewhat in virulence, and the amounts of the inoculum introduced were likely not entirely uniform—yet the probability of death throughout the sample was no more divergent than would be expected in groups randomly drawn from a population with constant probability of death.

This special method may be convenient if crossing over is suspected in genetic material which should theoretically segregate $r:1$. For the sake of easy comparison, let us apply it to the data of table 9.1, rearranged as in table 9.12. The new features are in the lower part of the table where chi-square is computed on two assumptions: (i) that the population ratio is 3:1, the value of chi-square being identical with the interaction chi-square calculated as a difference in section 9.2; and (ii) that there is no theoretical ratio. What are the advantages?

First, since $\chi^2 = 3.08$, $d.f. = 10$ is so small as to indicate no heteroge-

TABLE 9.12

SEGREGATION OF GREEN AND YELLOW SEEDLINGS IN 11 PROGENIES. DATA OF TABLE 9.1
Alternative method of computation

Number Examined	Number Yellow X	Probability of Yellow p	Products pX
122	24	0.1967	4.7208
149	39	0.2617	10.2063
86	18	0.2093	3.7674
55	13	0.2364	3.0732
71	17	0.2394	4.0698
179	38	0.2123	8.0674
150	30	0.2000	6.0000
36	9	0.2500	2.2500
91	21	0.2308	4.8468
53	14	0.2642	3.6988
111	26	0.2342	6.0892
1,103	$SX = 249$	$\bar{p} = 0.22575$ $\bar{q} = 0.77425$	$SpX = 56.7897$ $\bar{p}SX = 56.2118$ $SpX - \bar{p}SX = 0.5779$

Assuming 3:1 ratio, $p = 0.75$, $q = 0.25$

$$\chi^2 = \frac{SpX - \bar{p}SX}{pq} = \frac{0.5779}{(0.75)(0.25)} = 3.08$$

With no assumption about ratio,

$$\chi^2 = \frac{SpX - \bar{p}SX}{\bar{p}\bar{q}} = \frac{0.5779}{(0.2258)(0.7742)} = 3.31$$

neity among the sample ratios, the pooled chi-square is of interest and may conveniently be calculated by means of the formula (section 1.15),

$$\chi^2 = \frac{(a - rb)^2}{r(a + b)} = \frac{(854 - 3 \times 249)^2}{3(1,103)} = 3.46$$

We now have both the pooled and the interaction chi-squares which are essential: the total chi-square may be got by addition if it is needed. The calculations are notably less burdensome than those of table 9.1.

Second, we have as before the evidence to decide whether there is crossing over or not. While pooled chi-square is short of the 5% point, it does emphasize a bias—a deficit in the recessive yellows. If the investigator attributes this to crossing over (or to some other genetic phenomenon) he will reject the 3:1 hypothesis. He then has available the appropriate interaction $\chi^2 = 3.31$ based only on the assumption of independence, and this with a negligible amount of extra computation.

Fisher and Mather have made further adaptations of this method to test homogeneity in data with hierarchical subdivisions (8) (9).

EXAMPLE 9.15—In another sample of inoculated rats the data were: 25, 21; 50, 48; 20, 15; and 20, 17—the first number in each pair being the subsample size, the second, the number dying. Calculate $\chi^2 = 6.69$, $P = 8.6\%$.

EXAMPLE 9.16—Combine the values of chi-square for the 2 samples of rats, yielding $\chi^2 = 11.66$, $d.f. = 8$, $P = 17\%$.

EXAMPLE 9.17—If you are willing to accept the hypothesis that each sample of rats was drawn from a homogeneous population, it is appropriate to test the hypothesis that the probabilities of death in the two populations are the same; that is, that the two sample probabilities, 85.9% and 87.8%, differ only as may be expected in random sampling from a common population. Ans. Adjusted $\chi^2 = 0.070$ (section 9.7).

EXAMPLE 9.18—Apply the method of table 9.12 to the data of example 9.1, using percentages instead of probabilities. Note the slight modification in the formula.

Number in Group	Number Yellow X	Percentage Yellow p	
139	41	29.4964*	
102	31	30.3922	
179	52	29.0503	
86	25	29.0698	
122	36	29.5082	
628	$SX = 185$	$\bar{p} = 29.4586$ $\bar{q} = 70.5414$	$SpX = 5,451.17$ $\bar{p}SX = 5,449.84$
			$SpX - \bar{p}SX = 1.33$

For testing interaction from 3:1, $\chi^2 = \frac{100(SpX - \bar{p}SX)}{(25)(75)} = 0.071$

For testing independence, $\chi^2 = \frac{100(SpX - \bar{p}SX)}{\bar{p}\bar{q}} = 0.064$

* I carried 2 extra digits in order to verify the result in example 9.1.

Since we know from the earlier computations that pooled chi-square is significant at the 1% point, some such bias as crossing over may be decided upon. If so, the 3:1 hypothesis is abandoned and interaction calculated as 0.064. Clearly, interaction is negligible under either hypothesis—3:1 or independence.

EXAMPLE 9.19—Burnett (4) tried the effect of five storage locations on the viability of seed corn. In the kitchen garret, 111 kernels germinated among 120 tested; in a closed toolshed, 55 among 60; in an open toolshed, 55 among 60; out-of-doors, 41 among 48; and in a dry garret, 50 among 60. Calculate $\chi^2 = 5.09$, $d.f. = 4$, $P = 28\%$. The calculations are somewhat shorter if the numbers of nongerminating seeds are used. The results contain slight evidence against the hypothesis that there is a uniform probability of germination, irrespective of location.

EXAMPLE 9.20—Among some experiments reported by Lindstrom (14), the number of ears of corn displaying the presence of a recessive gene producing sugar was 18 among 33 having 8 rows of kernels, 37 among 63 with 10 rows, 27 among 70 with 12 rows, and none among 4 with 14 rows. It is easy to see that the expected number of sugary ears among the last four is less than five, the minimum expected number that should be used in applying chi-square. The 12-row and 14-row ears, may, therefore, be merged into one subsample having 27 sugary ears among 74. If this is done, $\chi^2 = 7.4$, $d.f. = 2$, and $P = 3\%$. The evidence is against homogeneity, though not strongly so. Apparently the sugary gene is associated with ears having few rows of kernels.

EXAMPLE 9.21—Using the data of the preceding example, calculate chi-square by means of the formula of section 9.9. Since 48.24% of the total 170 ears were sugary, one would expect 48.24% of the 33 eight-row ears to bear the gene; that is, 15.92 ears. The deviation from expected is $18 - 15.92 = 2.08$ ears. Similarly, five more deviations from expected are calculated (3 pairs in each of which the deviations are equal in absolute value but opposite in sign), the resulting chi-square being 7.4 as before.

9.11—An experiment requiring a test of technique. When novel experimental methods are used, the investigator must determine, among other things, whether he can reproduce his results; that is, whether he has adequate control over the conditions under which the experiment is performed. Reference to this problem was made in the section preceding. If results cannot be verified under controlled conditions assumed to be identical, then it is idle to try the effect of changing these conditions—one cannot know whether differences (or likenesses) in the results are to be charged to the controlled situation or to those unknown causes that elude control.

Decker and Andre (7) were faced with this problem when they started to investigate the effect of a short, sudden exposure to cold, the adults of the chinch bug being the organism studied. The experimental insects had to be gathered in the field, so that the degree of heterogeneity was unknown. Ten individuals were placed in each of 50 tubes and exposed for 15 minutes at -8°C . Since the chi-square method may not yield accurate results unless the expected numbers are all at least as great as five, the counts of the numbers dead in the individual tubes were combined at random in 5 lots of 10 tubes each; that is, lots of 100 nymphs. The mortalities were 14, 14, 23, 17, and 20 insects. From these data, $\chi^2 = 4.22$, $d.f. = 4$, and $P = 39\%$. This indicates about the variation that would be expected in sampling from homogeneous material. The mean probability of death was 17.6%. The results are in accord with the hypothesis that every individual was subject to this same chance of being killed by the exposure.

In a second sample of 500 adults, handled in the same manner except that they were exposed at -9°C ., the mortalities in groups of 100 were 38, 30, 30, 40, 27. The chi-square of 5.79 verifies the technique and shows only sampling variation from the estimated mortality of 33%. While no particular advantage is gained in this case, it is interesting to add the two values of chi-square, $\chi^2 = 4.22 + 5.79 = 10.01$, $d.f. = 8$, $P = 27\%$.

The gratifying uniformity in the experimental results leads one to place confidence in the rather surprising finding that the mortalities at -8°C . and -9°C . were significantly different. The total numbers dead in the two samples of 500 were 88 and 165. The value, $\chi^2 = 31.4$, $d.f. = 1$, P less than 0.0002 (section 8.8), constitutes convincing evidence that a rise in mortality with the lowering of temperature from -8°C . to -9°C . is a population characteristic, not merely an accident of sampling.

The ease of applying such a test of experimental technique makes its

use almost a routine procedure except in highly standardized processes. It is necessary merely to collect the data in several small groups instead of in one mass. The additional information available may modify conclusions profoundly.

EXAMPLE 9.22—Dr. R. H. Walker of the Intermountain Forest and Range Experiment Station has provided the following unpublished data on counts of bacteria in suspensions of soil and indigo particles, following the method of Thornton and Gray (19). The number of indigo particles in a field specifies the amount of soil present. Good technique will be indicated if the numbers of bacteria vary from proportionality to the indigo particles only so much as would be expected in random sampling from a population in which this ratio is constant. On a slide, 16 fields were counted, four in each of four drops. The data are records of the total slide counts, four slides to each soil suspension. For purposes of computation, we shall give only two figures: (i) the total count of both bacteria and indigo particles, and (ii) the number of bacteria. For the four plates of soil suspension 1: 204, 78; 260, 75; 246, 76; 278, 95. Chi-square = 5.22, *d.f.* = 3, *P* = 16%. For suspension 2: 260, 60; 196, 50; 198, 45; 186, 50. Chi-square = 1.30, *d.f.* = 3, *P* = 73%. For suspension 3: 177, 22; 177, 23; 150, 16; 177, 20. Chi-square = 0.53, *d.f.* = 3, *P* = 91%. For suspension 4: 289, 46; 356, 63; 281, 45; 250, 42. Chi-square = 0.48, *d.f.* = 3, *P* = 92%. The agreement is closer than one would ordinarily expect in random sampling.

EXAMPLE 9.23—To determine whether the four soil suspensions differed significantly in the numbers of bacteria per unit volume, the totals are taken from the records of the foregoing example: for suspension 1: 988, 324; 2: 840, 205; 3: 681, 81; 4: 1176, 196. Chi-square = 130, *d.f.* = 3, *P* less than 1%. There is no doubt of differences among the four soil suspensions.

EXAMPLE 9.24—In computing chi-square by the method of table 9.11, the numbers of observations in the groups may be either equal or different. If they are equal, the mathematician can easily reduce the formula to the special one $\chi^2 = kSx^2/\bar{x}(k - \bar{x})$, where *k* is the size of the subsample, Sx^2 is the sum of the squares of the deviations from mean, $\bar{x}$, obtained in the usual manner by using the formula, $Sx^2 = SX^2 - (SX)^2/n$. Those who work extensively with chi-square will find some saving in time by using this special formula when their group numbers are equal. They may practice by applying it to the Decker and Andre data. This formula was first used by "Mathetes" (15).

9.12—An experiment in which the presence of an attribute is recorded quantitatively. In the enumeration data heretofore considered, the mere presence of some attribute was recorded. There is another kind in which presence may be observed in more than one unit, so that the record becomes quantitative even at this initial stage. Both kinds are employed in many studies of insect populations. In investigations of the intensity of attack by codling moth, for example, the fruits may be designated qualitatively as merely injured or free; or they may be described as containing 0, 1, 2, etc., larvae. The latter kind of data often follows a distribution known as the Poisson, further account of which will be given in chapter 16. An example is found in an experiment designed to investigate various treatments for the control of cabbage loopers (5). Certain parts of the data are summarized in table 9.13. The original record of the counts on the individual plants in plot 2 of treatment four was like this: 0, 1, 0, 0, 1, 2, 1, 0, 0, 0, 0, 4, 0, etc. In the four plots receiving one of the treatments not entered in the table only a single looper was found. For this illustration we selected only those treatments yielding

TABLE 9.13
NUMBER OF LOOPERS OBSERVED ON 50 CABBAGE PLANTS IN A PLOT
Four plots treated alike. Five treatments

Treatment	Number of Loopers in Each of 4 Plots X	Total for Treatment SX	Mean $\bar{x}$	Chi-square $Sx^2/\bar{x}$
1	11, 4, 4, 5	24	6.00	5.67
2	6, 4, 3, 6	19	4.75	1.42
3	8, 6, 4, 11	29	7.25	3.69
4	14, 27, 8, 18	67	16.75	11.39
5	7, 4, 9, 14	34	8.50	6.24
Total		173		28.41

Computation for treatment 1

$$Sx^2 = SX^2 - (SX)^2/n = (11)^2 + (4)^2 + (4)^2 + (5)^2 - (24)^2/4 = 34$$

$$\chi^2 = Sx^2/\bar{x} = 34/6.00 = 5.67, d.f. = 3, P = 14\%$$

Computation for Five Treatments

$$Sx^2 = SX^2 - (SX)^2/n = (24)^2 + (19)^2 + (29)^2 + (67)^2 + (34)^2 - (173)^2/5 = 1437.2$$

$$\bar{x} = 173/5 = 34.6$$

$$\chi^2 = Sx^2/\bar{x} = 1437.2/34.6 = 41.5, d.f. = 4, P \text{ less than } 0.0002$$

at least five as the expected number for each plot. On plots with less loopers there was even greater uniformity than on those selected for this investigation.

For treatment 1, the expected number in each plot is simply the mean of the four plot counts, 6 loopers. The deviations from expected are, therefore, just the same as deviations from mean in measurement data, $11 - 6$, $4 - 6$, etc. To make up chi-square, each of these deviations must be squared, then divided by the expected number. Since the latter is the same for the four plots, division may be delayed until the addition has been completed. Thus,

$$\chi^2 = \frac{(5)^2 + (-2)^2 + (-2)^2 + (-1)^2}{6} = \frac{34}{6} = 5.67$$

In the table, the computation is done by the usual machine method. It is easy to see that the formula used is merely an adaptation of the customary formula for chi-square.

For treatment 1, the value of chi-square, computed in the table, would be exceeded 14 times in 100 trials from homogeneous material. The uniformity of infestation need not be questioned. The sum of the chi-squares for the five treatments is 28.41, $d.f. = 15$, $P = 2\%$. This indicates significant variation from uniformity of infestation among plots treated alike. From the original records, this lack of homogeneity seemed to be confined chiefly to the plots more severely attacked. However, the departure from uniformity appears to be slight compared with the varia-

tion induced by the treatments. This is shown by the last computation in the table where five treatment totals are examined. The chi-square of 41.5, $d.f. = 4$, leaves little doubt of the treatment effects, even though an unknown amount of variation may have been introduced by the heterogeneity of the plots treated alike.

EXAMPLE 9.25—Davis (6) counted the number of nodal smut galls in the corn plants of eight plots. The plants of four plots had been inoculated with the organism producing the galls. The number of galls observed in these plots were 56, 60, 41, and 75. Chi-square = 10.1, $d.f. = 3$, $P = 2\%$. In the four plots whose plants were not inoculated, the numbers of galls were 20, 12, 26, 22. Chi-square = 5.2, $d.f. = 3$, $P = 17\%$. For the two groups, $\chi^2 = 15.3$, $d.f. = 6$, $P = 2\%$. These figures indicate a significant lack of uniformity among the plots treated alike. Nevertheless, since the heterogeneity is not great, it is proper to test the difference between the totals for the two groups of plots. For the sums, 232 and 80, $\chi^2 = 74$, $d.f. = 1$, P less than 1%. There is little question of a population difference between the totals.

EXAMPLE 9.26—In example 9.22, soil suspension 1, ignore the numbers of indigo particles and compute, for the bacteria alone, $\chi^2 = 3.3$, $d.f. = 3$, $P = 36\%$. Neither the bacteria nor the ratios of bacteria to totals depart significantly from sampling variation in a homogeneous population. The ratios, as shown by the larger chi-square, are more variable, possibly because the test is more critical; the former test was based on the variation of both bacteria and indigo particles.

EXAMPLE 9.27—Davis (6) counted the numbers of ears on two plots of corn whose plants had been inoculated with an organism producing smut, and also on two check plots. The numbers were 436 and 432, respectively. The obvious failure of the inoculation to affect the number of ears is verified by unadjusted $\chi^2 = 0.02$, $d.f. = 1$. The numbers of smutted ears, however, were 69 and 39. That the percentages differ significantly is shown by adjusted $\chi^2 = 8.59$, $d.f. = 1$.

9.13—Summary. The methods of chapter 1 have been extended to more than 2 classes (e.g., the 9:3:3:1 genetic ratio, the 2×2 and the $2 \times 2 \times 2$ tables) and to 2 or more samples. These samples may be from populations with or without theoretical ratios. The Poisson distribution has also been introduced.

If 2 or more samples are drawn from populations with theoretical ratios, 3 different hypotheses may be tested: (i) that the pooled sample is from the specified population, (ii) that the several samples may all have come from the specified population, and (iii) that there is no interaction.

In the 2×2 table, the hypothesis tested by chi-square may be stated in 3 ways: (i) the two attributes are independent. (ii) there is no interaction in the population; that is, $ad = bc$. (iii) pairs of ratios such as $a/(a + b)$ and $c/(c + d)$ are random samples from a common population ratio—the hypothesis of homogeneity. In the $2 \times 2 \times 2$ table, the chi-square test of independence also tests the hypothesis of no second order interaction, $adb'c' = a'd'bc$.

In the $R \times C$ table, the hypothesis tested by chi-square is that of independence (or homogeneity) or that of absence of interaction.

A special method is given for computing chi-square in an $R \times 2$ table of probabilities or percentages. At the end, there are available chi-squares for testing interaction under either of 2 hypotheses, (i) that the

population has a theoretical ratio or (ii) that the attributes are independent.

Computing methods are described for samples from Poisson distributions.

REFERENCES

1. M. S. BARTLETT. Supplement to the Journal of the Royal Statistical Society, 2:248 (1935).
2. W. BATESON and R. C. PUNNETT. Journal of Genetics, 1:297 (1911).
3. MARK F. BOYD, S. F. KITCHEN and HUGO MUENCH. The American Journal of Tropical Medicine, 16:589 (1936).
4. L. C. BURNETT. Thesis submitted for the degree Master of Science, Iowa State College (1906).
5. D. J. CAFFREY and C. E. SMITH. Bureau of Entomology and Plant Quarantine, USDA. Data from Baton Rouge (1934).
6. GLEN N. DAVIS. Iowa Agricultural Experiment Station Research Bulletin 199 (1936).
7. GEORGE C. DECKER and FLOYD ANDRE. Iowa State College Journal of Science, 10:403 (1936).
8. R. A. FISHER. Statistical Methods for Research Workers. Oliver and Boyd, Edinburgh (1934-1944).
9. R. A. FISHER and K. MATHER. Annals of Eugenics, 7:265 (1936).
10. C. H. GOULDEN. Methods of Statistical Analysis. John Wiley and Sons, New York (1939).
11. J. W. GOWEN. Unpublished data, Iowa Agricultural Experiment Station (1945).
12. T. ROY HANSBERRY and CHARLES H. RICHARDSON. Iowa State College Journal of Science, 10:27 (1935).
13. E. W. LINDSTROM. Cornell University Agricultural Experiment Station Memoir, 13 (1918).
14. E. W. LINDSTROM. Iowa Agricultural Experiment Station Research Bulletin 142 (1931).
15. "Mathetes." Annals of Applied Biology, 11:220 (1924).
16. ERNST MAYR and TH. DOBZHANSKY. Proceedings of the National Academy of Science, 31:75 (1945).
17. GEORGE W. SNEDECOR and M. R. IRWIN. Iowa State College Journal of Science, 8:75 (1933).
18. NORMAN V. STRAND and RAYMOND J. JESSEN. Iowa Agricultural Experiment Station Research Bulletin 315 (1943).
19. H. P. THORNTON and P. H. H. GRAY. Proceedings of the Royal Society of London, Series B, 115:522 (1934).
20. RAY E. WAKELEY. Iowa Agricultural Experiment Station Research Bulletin 249 (1938).
21. F. YATES. Journal of the Royal Statistical Society Supplement, 1:217 (1934).

Chapter 10

TWO OR MORE GROUPS OF MEASUREMENT DATA. ANALYSIS OF VARIANCE

10.1—Extension from two groups to many. Statistical and experimental methods for two groups were presented in chapter 4, but the needs of the investigator are seldom confined to only two treatments. More than two require extensions of the methods. For enumeration data such extensions were made in the foregoing chapter. We are now ready to do the same for measurement data, expand chapter 4 to allow for the simultaneous trial of three or more treatments.

During the early 1920's great progress was made in handling the statistics of multiple groups (2). Near the beginning of the decade Fisher (3) was able to solve some problems of distribution that completed his *analysis of variance*. This device has led to tremendous expansion in the design of experiments, each with an appropriate statistical analysis. Some of the simplest of these form the topic of the present chapter.

10.2—Multiple samples from a common population. Analysis of variance. In the earlier sections of chapter 4 we considered two groups of equal size: three or more such groups now concern us. As examples, there may be several lots of animals, every lot receiving a different ration; or several classes of children in the 6th grade on which different methods of instruction are being tried. The statistical analysis of the ensuing experimental data is most easily explained by reference to the sampling of chapter 3. There we drew at random a number of groups from a normally distributed set of pig gains with mean, 30 pounds, and variance, 100. If we put together several of these groups, say 4, we shall have in effect an experiment with 40 pigs divided into 4 lots of 10 animals each. Ordinarily the lots would receive different treatments, but our data simulate a *dummy experiment*, or *uniformity trial*, in which all lots are treated alike.

The results of an experiment of the kind just described are recorded in table 10.1. For ease in calculation the lots have been reduced to 5 pigs each and the gains in weight changed slightly so as to avoid fractions. It is convenient to refer to the entire experiment as a sample: this is divided into n subsamples (lots or groups) each with k items. Thus, the experiment is comprised of nk individuals.

The entire sample of 20 weight gains has the mean 29 pounds, an estimate of $m = 30$ pounds. The variance, $1,918/19 = 100.9$, estimates $\sigma^2 = 100$. These statistics could be used, if desired, to set fiducial limits or to test hypotheses in the manner of chapter 2.

TABLE 10.1
GAINS IN WEIGHT (POUNDS) OF 4 LOTS OF SWINE, 5 PER LOT. SAMPLE FROM TABLE 3.1

$k = 5$ Pigs per Lot	$n = 4$ Lots				Entire Sample
	1	2	3	4	
1	40	29	11	17	
2	24	27	31	21	
3	46	20	17	28	
4	20	39	37	33	
5	35	45	39	21	
SX	165	160	135	120	580
$\bar{x}$	33	32	27	24	29
SX^2	5,917	5,516	4,261	3,044	18,738
$(SX)^2/k$	5,445	5,120	3,645	2,880	16,820
$*Sx^2$	472	396	616	164	1,918

* This method of calculation is explained in section 5.2.

A second estimate of the population variance is got from the sums of squares within the 4 subsamples. Following the method of chapter 4, these sums of squares are added or pooled:

$$472 + 396 + 616 + 164 = 1,648$$

What about degrees of freedom? Since the mean of each lot is used to measure deviations from, one degree of freedom is deducted from the original 5. Altogether, then, there are 4 *d.f.* in each of the 4 subsamples, these 16 corresponding to the sum of squares, 1,648. The resulting mean square, $1,648/16 = 103$, is the second estimate of σ^2 furnished by the sample.

The lot means lead to a third estimate of the population variance. Because these 4 means are random samples from the same normal population, their variance is an unbiased estimate of $\sigma^2/k = 100/5 = 20$. The variance of these means is

$$\frac{(33 - 29)^2 + (32 - 29)^2 + (27 - 29)^2 + (24 - 29)^2}{3} = 18$$

instead of 20. Since 18 is an estimate of $\sigma^2/5$, then $5 \times 18 = 90$ is the third estimate of σ^2 , this one based on the subsample means with 3 degrees of freedom.

The foregoing analysis is summarized in table 10.2. One additional feature: the variance of lot means has been multiplied by its degrees of freedom to form a sum of squares, 270, the utility of which will be clear from what follows.

The table brings into prominence two notable facts: not only the total sum of squares but also the corresponding degrees of freedom have been

TABLE 10.2
ANALYSIS OF VARIANCE OF PIG GAINS IN TABLE 10.1

Source of Variation	Degrees of Freedom	Sum of Squares	Estimates of σ^2
Individuals of the several lots	16	1,648	103
Lot Means	3	270	90
Total	19	1,918	100.9

separated into two parts, both associated with the structure of the experiment. Since the parts as well as the total were computed independently, their additive relationship is emphasized. You will recall similar addition theorems in section 6.12. This partition of degrees of freedom and corresponding sums of squares, each part providing an estimate of variance, is called *analysis of variance*.

EXAMPLE 10.1—Analyze the variance in the following data selected for ease in computation.

Observation Number	Subsample Number			
	1	2	3	4
1	11	13	21	10
2	4	9	18	4
3	6	14	15	19

Analysis of Variance

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Individuals	8	172	21.5
Subsamples	3	186	62.0
Total	11	358	32.5

EXAMPLE 10.2—Calculate the sum of squares for subsamples, 186, by adding the four subsample correction terms, 147, 432, etc., then deducting the sample correction term, 1,728.

EXAMPLE 10.3—Calculate the sum of squares for subsamples as follows:

$$\frac{(21)^2 + (36)^2 + (54)^2 + (33)^2}{3} - \frac{(144)^2}{12}$$

The numbers squared in the numerator are the subsample and sample totals, while those in the denominators are k and nk .

Which method of calculation seems easiest?

10.3—Calculation of analysis of variance. Samples from different populations. One of the reasons for the popularity of analysis of variance is the elegance of its computations. Those in table 10.1 may be shortened considerably as soon as their meaning is learned. As a preliminary, note that, in the last line but one, the sum of the four correction terms, corrected by the one in the last column, is equal to the sum of squares for subsample means:

$$5,445 + 5,120 + 3,645 + 2,880 - 16,820 = 270$$

This obviates the separate calculation of the variance of the means and its subsequent multiplication by k and $n - 1$.

With adequate facilities for checking calculations, most people prefer to omit the last three lines of the central section of table 10.1, calculating the two sums of squares, "total" and "lot means," then subtracting for "individuals." This is feasible because the sums, SX , instead of the correction terms, $(SX)^2/k$, can be used in calculating the sum of squares for lot means, thus:

$$\frac{(165)^2 + (160)^2 + (135)^2 + (120)^2}{5} - \frac{(580)^2}{20} = 270$$

The only change is that division by k takes place after the squares are added instead of before. The form is easily remembered because the denominator in each part is the number of individuals, k or nk , the square of whose sum is in the numerator.

To illustrate this abbreviated calculation, and also to introduce an experiment in which the subsamples may not be drawn from a common population, we use the data in table 10.3. During cooking, doughnuts absorb fat in various amounts. Lowe (9) wished to learn if the amount absorbed is characteristic of the fat used. Six mixes, each containing 24 doughnuts, were cooked in each of 8 fats. The numbers in the table are the grams of fat absorbed per mix. The mix means, unnecessary in the computations, should always be recorded in the table.

The two necessary parts of the computation are (i) the *total* sum of squares, 9,193, in the righthand column, calculated by summing the X^2 's over the whole table, ignoring groups, then correcting for mean in the usual manner; and (ii) the sum of squares of subclass (fat) means, 3,527, got from the subclass sums as described above. The remainder of the calculations follow immediately from these two sums of squares.

The beginner would do well to sum the squares, SX^2 , in each subclass separately, then add these 8 sums to verify that at the right, 262,364.

A symbolical summary of the calculations, table 10.4, may be helpful to some readers.

If no calculating machine is available, proceed as follows: (i) Code each subclass separately by subtracting a convenient number about the size of its mean, then calculate its Sx^2 . The sum of n such is the sum of

TABLE 10.3
GRAMS OF FAT ABSORBED BY MIXES OF 24 DOUGHNUTS

Row No.	Fat Number								Total
	1	2	3	4	5	6	7	8	
1	164	173	177	178	163	175	178	155	
2	172	161	183	191	165	193	146	166	
3	168	190	197	197	144	178	141	149	
4	177	180	169	182	177	171	150	164	
5	156	197	179	185	165	163	169	170	
6	195	167	187	177	176	176	182	168	
*SX	432	468	492	510	390	456	366	372	3,486
*SX ²									262,364
*C = (SX) ² /nk									253,171
Sx ²									9,193
$\bar{x}$	172	178	182	185	165	176	161	162	172.6

Fats:

$$\frac{(432)^2 + (468)^2 + \dots + (372)^2}{6} - 253,171 = \frac{1,540,188}{6} - 253,171 = 3,527$$

Analysis of Variance

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Total	47	9,193	
Fats	7	3,527	503.9
Mixes within fats	40	5,666	141.6

* In the calculation of these 3 lines, the digit 1 in the first place is ignored; that is, the data are coded by subtracting 100.

squares for mixes, 5,666. (ii) Code the 8 subclass means by subtracting a suitable number, say 170, calculate their sum of squares, then multiply by k . This yields the sum of squares for subclasses, 3,527. If desired, the *total* sum of squares may be got by adding the other two, but there is little occasion for computing it.

10.4—Test of a null hypothesis. The variance ratio, F. The doughnut data of table 10.3 introduce a familiar question: is ordinary random sampling accountable for the large discrepancy between the mean squares for fats and mixes, or shall we conclude that the fat means are differentiated by causes other than sampling fluctuations? Lowe suspected that the nature of the several fats might result in different amounts of absorption. If so, the mixes are samples from different populations instead of a common one. Even if these populations are all assumed to be normally

TABLE 10.4
 SYMBOLICAL SUMMARY OF COMPUTATIONS IN ANALYSIS OF VARIANCE
 Single criterion of classification; n subsamples, each with k items.
 Total number of items in sample = nk

The data						
	Group 1	Group 2	...	Group p	...	Group n
	X_1	X_2		X_p		X_n
	$\vdots$	$\vdots$		$\vdots$		$\vdots$
	$\vdots$	$\vdots$		$\vdots$		$\vdots$
	k items	k items		k items		k items
Sums:	S_1	S_2	...	S_p	...	S_n
The computations						
1. The sum of all items: $S = S_1 + S_2 + \dots + S_n$						
2. The correction term: $C = S^2/nk$						
3. Total sum of squares: $SX^2 - C$						
4. Groups: $(S_1^2 + S_2^2 + \dots + S_n^2)/k - C$						
5. Individuals: total - groups						
Degrees of freedom						
Total		$nk - 1$				
Groups		$n - 1$				
Individuals		$nk - n = n(k - 1)$				

distributed, they may have different means or different variances or both.

One null hypothesis that can be tested is that all the mixes are random samples from the same normal population, as were the lots of table 10.1. For evidence about ordinary sampling variation under this hypothesis, we go back to the drawings from table 3.1. Suppose the original groups of 10 gains are now combined into samples of 10 groups as in table 10.5. The analysis of variance of the entire 100 items shows that the ratio of the two independent estimates of variance, group means and individuals, is $172.3/114.0 = 1.51$ and we know that sampling variation accounts for the failure of these two estimates to be equal. It is necessary next to learn the magnitude of variation ordinarily encountered in such ratios.

The ratio of the two estimates of variance,

$$\frac{\text{Mean square of group means}}{\text{Mean square of individuals}},$$

has a distribution discovered by Fisher. I named it F in his honor (14). Fisher tabulated the distribution in the form, $z = \log_e \sqrt{F}$ (4). Fisher and Yates (5) designate F as the *variance ratio*, while Mahalanobis (10), who first calculated it, called it x .

The distribution of F in 100 samples like that of table 10.5 is shown in table 10.6. One notices first the skewness: a concentration of small

TABLE 10.5

TEN GROUPS, EACH WITH 10 ITEMS, RANDOMLY DRAWN FROM THE NORMAL DISTRIBUTION OF TABLE 3.1. THIS SIMULATES A DUMMY EXPERIMENT WITH 10 LOTS OF PIGS

1	2	3	4	5	6	7	8	9	10
35	34	17	39	18	7	33	42	7	39
42	38	29	34	7	22	29	35	48	34
42	26	30	22	31	40	25	31	43	33
30	17	36	27	41	29	27	33	53	33
15	42	41	42	21	31	21	53	7	33
31	28	30	33	17	41	46	21	33	39
29	35	3	24	21	30	19	41	17	36
29	33	23	36	40	12	43	29	57	32
17	16	38	29	14	44	22	34	42	32
21	40	30	25	14	30	21	49	42	30

Analysis of Variance

Source of Variation	Degrees of Freedom	Mean Square
Group means	9	172.3
Individuals	90	114.0

$$F = 172.3/114.0 = 1.51$$

values and a long tail of larger. Next, there is the fact that 65 of the F 's are less than 1. If you remember that both terms of the ratio are estimates of σ^2 , you may be surprised that 1 is not the median. However, the mean, calculated as in section 8.3, is 0.96, about the average one would expect. Finally it is to be observed that 5% of the values lie beyond 2.25 and 1% beyond 2.75, so that these points are roughly comparable with those of the theoretical distribution next to be considered.

Table 10.7 contains the theoretical 5% and 1% points of F for convenient combinations of degrees of freedom. Across the top of the table is found n_1 d.f. corresponding to the number of subsamples: $n_1 = n - 1$. At the left is n_2 , the degrees of freedom for individuals, $n(k - 1)$. To find the 5% and 1% points for table 10.5, look in the column headed by

TABLE 10.6

DISTRIBUTION OF F IN 100 SAMPLES FROM TABLE 3.1. DEGREES OF FREEDOM 9 AND 90

Class Interval	Frequency	Class Interval	Frequency
0. -0.24	7	1.50-1.74	5
0.25-0.49	16	1.75-1.99	2
0.50-0.74	16	2.00-2.24	4
0.75-0.99	26	2.25-2.49	2
1.00-1.24	11	2.50-2.74	2
1.25-1.49	8	2.75-2.99	1

$n_1 = 9$ and down to the row $n_2 = 90$. The points are 1.98 and 2.62, halfway between the values for 80 *d.f.* and 100. In order to check the sampling distribution more exactly, I went back to the original calculations and found 8% of our F 's beyond the 5% point and 2% beyond the 1%. This gives some idea of the variation to be encountered.

For the doughnut experiment, the hypothesis set up—that the mix data are random samples from a common population—may be judged by means of table 10.7. From the analysis of variance in table 10.3,

$$F = 503.9/141.6 = 3.56$$

For $n_1 = 7$ and $n_2 = 40$ the 1% point in the table is 3.12. Thus there is less than one chance in 100 of drawing a sample having a larger value of F from the distribution specified in the hypothesis. Evidently the subsamples come from different populations.

The conclusion just reached may mean that each of the 8 fats constitutes a separate population of fat absorption, or it may mean that some of the fats belong to one population and some to another. There might be any number of populations from 2 to 8. The analysis of variance does not provide information on this point: there is no way of distinguishing the subsamples of one population from those of another. However, the fat means, each an unbiased estimate of its population mean, contain available evidence. Fat 4 ranked highest in absorption by the doughnuts, followed closely by 3. Fats 7, 8, and 5 were lowest. Further evidence is found in the fiducial interval. If 141.6 be accepted as an estimate of a common variance in all the populations, with 40 degrees of freedom, the 95% semi-interval is $t_{.05, 40} \sqrt{141.6/6} = 10$ pounds (9.82, more exactly). Thus, m for fat 4 may be around $185 - 10 = 175$ pounds, and that for fat 7 may be near $161 + 10 = 171$ pounds. You see, it is unlikely that these two fats have the same population mean. As for fats 3 and 8, their m 's might both be 172, but such a coincidence is improbable. Now, this is about all that can be learned from the fiducial interval, because fats 2 and 5 may well have the same m . It seems fairly clear that fats 3 and 4 may belong to one population while 7 and 8 belong to another. There may be a population with an intermediate mean, and this may include fats 1, 2, and 6. What the experiment tells us certainly is (i) that the 8 subsamples are probably not drawn from a common population, and (ii) that whatever population means are involved are estimated in an unbiased manner by the subsample means.

As you progress in your study of experimental design you will find that experiments can be set up to test simultaneously a whole array of null hypotheses. The simple structure of the doughnut experiment cannot be expected to support more than the single test described.

Three remarks close this section. (i) The F distribution is based on random sampling from normally distributed populations, but it has been found (12) that little bias is introduced into the test of significance by

TABLE 10.7—5% (ROMAN TYPE) AND 1% (BOLD FACE TYPE) POINTS FOR THE DISTRIBUTION OF F

No.	n, degrees of freedom (for greater mean square)																								No.
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	CO	
1	161	200	216	225	230	234	237	239	241	242	243	244	245	246	248	249	250	251	252	253	253	254	254	254	254
2	4,852	4,999	5,463	5,625	5,764	5,859	5,928	5,981	6,022	6,056	6,082	6,106	6,122	6,139	6,208	6,234	6,258	6,286	6,302	6,323	6,334	6,352	6,361	6,366	
3	18.51	19.00	19.16	19.25	19.30	19.33	19.36	19.37	19.38	19.39	19.40	19.41	19.42	19.43	19.44	19.45	19.46	19.47	19.47	19.48	19.49	19.49	19.50	19.50	
4	98.49	99.00	99.17	99.25	99.30	99.33	99.34	99.36	99.38	99.40	99.41	99.42	99.43	99.44	99.45	99.46	99.47	99.48	99.48	99.49	99.49	99.49	99.50	99.50	
5	10.13	9.55	9.28	9.12	9.01	8.94	8.88	8.84	8.81	8.78	8.76	8.74	8.71	8.69	8.66	8.64	8.62	8.60	8.58	8.57	8.56	8.54	8.54	8.53	
6	34.12	34.82	29.46	28.71	28.24	27.91	27.67	27.49	27.34	27.23	27.13	27.05	26.92	26.83	26.69	26.60	26.50	26.41	26.35	26.27	26.23	26.18	26.14	26.12	
7	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.93	5.91	5.87	5.84	5.80	5.77	5.74	5.71	5.70	5.68	5.66	5.65	5.64	5.63	
8	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.54	14.45	14.37	14.24	14.15	14.02	13.93	13.83	13.74	13.69	13.61	13.57	13.52	13.48	13.46	
9	6.01	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.78	4.74	4.70	4.68	4.64	4.60	4.56	4.53	4.50	4.46	4.44	4.42	4.40	4.38	4.37	4.36	
10	16.26	13.27	12.06	11.39	10.97	10.67	10.45	10.27	10.15	10.05	9.96	9.89	9.77	9.68	9.55	9.47	9.38	9.29	9.24	9.17	9.13	9.07	9.04	9.02	
11	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00	3.96	3.92	3.87	3.84	3.81	3.77	3.75	3.72	3.71	3.69	3.68	3.67	
12	13.74	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.79	7.72	7.60	7.52	7.39	7.31	7.23	7.14	7.09	7.02	6.99	6.94	6.90	6.88	
13	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.63	3.60	3.57	3.52	3.49	3.44	3.41	3.38	3.34	3.32	3.29	3.28	3.25	3.24	3.23	
14	12.25	9.55	8.45	7.85	7.46	7.19	7.00	6.84	6.71	6.62	6.54	6.47	6.35	6.27	6.15	6.07	5.98	5.90	5.85	5.78	5.75	5.70	5.67	5.65	
15	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.34	3.31	3.28	3.23	3.20	3.15	3.12	3.08	3.05	3.03	3.00	2.98	2.96	2.94	2.93	
16	11.26	8.65	7.59	7.01	6.63	6.37	6.19	6.03	5.91	5.82	5.74	5.67	5.56	5.48	5.36	5.28	5.20	5.11	5.06	5.00	4.96	4.91	4.88	4.86	
17	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.13	3.10	3.07	3.02	2.98	2.93	2.90	2.86	2.82	2.80	2.77	2.76	2.73	2.72	2.71	
18	10.56	8.02	6.99	6.42	6.06	5.80	5.62	5.47	5.35	5.26	5.18	5.11	5.00	4.92	4.80	4.73	4.64	4.56	4.51	4.45	4.41	4.36	4.33	4.31	
19	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.97	2.94	2.91	2.86	2.82	2.77	2.74	2.70	2.67	2.64	2.61	2.59	2.56	2.55	2.54	
20	10.04	7.56	6.55	5.99	5.64	5.39	5.21	5.06	4.95	4.85	4.78	4.71	4.60	4.52	4.41	4.33	4.25	4.17	4.12	4.05	4.01	3.96	3.93	3.91	
21	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.86	2.82	2.79	2.74	2.70	2.65	2.61	2.57	2.53	2.50	2.47	2.45	2.42	2.41	2.40	
22	9.65	7.20	6.22	5.67	5.32	5.07	4.88	4.74	4.63	4.54	4.46	4.40	4.29	4.21	4.10	4.02	3.94	3.86	3.80	3.74	3.70	3.66	3.62	3.60	
23	4.75	3.88	3.49	3.26	3.11	3.00	2.92	2.85	2.80	2.76	2.72	2.69	2.64	2.60	2.54	2.50	2.46	2.42	2.40	2.38	2.35	2.32	2.31	2.30	
24	9.33	6.93	5.95	5.41	5.06	4.82	4.65	4.50	4.39	4.30	4.22	4.16	4.05	3.98	3.86	3.78	3.70	3.61	3.56	3.49	3.46	3.41	3.38	3.36	
25	4.67	3.80	3.41	3.18	3.02	2.92	2.84	2.77	2.72	2.67	2.63	2.60	2.55	2.51	2.46	2.42	2.38	2.34	2.32	2.28	2.26	2.24	2.22	2.21	
26	9.07	6.70	5.74	5.20	4.86	4.62	4.44	4.30	4.19	4.10	4.02	3.96	3.85	3.78	3.67	3.59	3.51	3.42	3.37	3.30	3.27	3.21	3.18	3.16	

TABLE 10.7-5% (ROMAN TYPE) AND 1% (BOLD FACE TYPE) POINTS FOR THE DISTRIBUTION OF F

n ₁	n ₂ degrees of freedom (for greater mean square)																						n ₁		
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200		500	∞
14	4.60	3.74	3.34	3.11	2.96	2.85	2.77	2.70	2.65	2.60	2.56	2.53	2.48	2.44	2.39	2.35	2.31	2.27	2.24	2.21	2.19	2.16	2.16	2.14	2.13
15	8.86	6.51	5.56	5.03	4.69	4.46	4.28	4.14	4.03	3.94	3.86	3.80	3.70	3.62	3.51	3.43	3.34	3.26	3.21	3.14	3.11	3.06	3.02	3.00	
15	4.54	3.68	3.29	3.06	2.90	2.79	2.70	2.64	2.59	2.55	2.51	2.48	2.43	2.39	2.33	2.29	2.25	2.21	2.18	2.15	2.12	2.10	2.08	2.07	
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.73	3.67	3.56	3.48	3.36	3.29	3.20	3.12	3.07	3.00	2.97	2.92	2.89	2.87	
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.45	2.42	2.37	2.33	2.28	2.24	2.20	2.16	2.13	2.09	2.07	2.04	2.02	2.01	
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.61	3.55	3.45	3.37	3.25	3.18	3.10	3.01	2.96	2.89	2.86	2.80	2.77	2.75	
17	4.45	3.59	3.20	2.96	2.81	2.70	2.62	2.55	2.50	2.45	2.41	2.38	2.33	2.29	2.23	2.19	2.15	2.11	2.08	2.04	2.02	1.99	1.97	1.96	
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.52	3.45	3.35	3.27	3.16	3.08	3.00	2.92	2.86	2.79	2.76	2.70	2.67	2.65	
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.37	2.34	2.29	2.25	2.19	2.15	2.11	2.07	2.04	2.00	1.98	1.95	1.93	1.92	
18	8.28	6.01	5.09	4.58	4.25	4.01	3.85	3.71	3.60	3.51	3.44	3.37	3.27	3.19	3.07	3.00	2.91	2.83	2.78	2.71	2.68	2.62	2.59	2.57	
19	4.38	3.52	3.13	2.90	2.74	2.63	2.55	2.48	2.43	2.38	2.34	2.31	2.26	2.21	2.15	2.11	2.07	2.02	2.00	1.96	1.94	1.91	1.90	1.88	
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.36	3.30	3.19	3.12	3.00	2.92	2.84	2.76	2.70	2.63	2.60	2.54	2.51	2.49	
20	4.35	3.49	3.10	2.87	2.71	2.60	2.52	2.45	2.40	2.35	2.31	2.28	2.23	2.18	2.12	2.08	2.04	1.99	1.96	1.92	1.90	1.87	1.85	1.84	
20	8.10	5.85	4.94	4.43	4.10	3.87	3.71	3.56	3.45	3.37	3.30	3.23	3.13	3.05	2.94	2.86	2.77	2.69	2.63	2.56	2.53	2.47	2.44	2.42	
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.28	2.25	2.20	2.15	2.09	2.05	2.00	1.96	1.93	1.89	1.87	1.84	1.82	1.81	
21	8.02	5.78	4.87	4.37	4.04	3.81	3.65	3.51	3.40	3.31	3.24	3.17	3.07	2.99	2.88	2.80	2.72	2.63	2.58	2.51	2.47	2.42	2.38	2.36	
22	4.30	3.44	3.05	2.82	2.66	2.55	2.47	2.40	2.35	2.30	2.26	2.23	2.18	2.13	2.07	2.03	1.98	1.93	1.91	1.87	1.84	1.81	1.80	1.78	
22	7.94	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.18	3.12	3.02	2.94	2.83	2.75	2.67	2.58	2.53	2.46	2.42	2.37	2.33	2.31	
23	4.28	3.42	3.03	2.80	2.64	2.53	2.45	2.38	2.32	2.28	2.24	2.20	2.14	2.10	2.04	2.00	1.96	1.91	1.88	1.84	1.82	1.79	1.77	1.76	
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.14	3.07	2.97	2.89	2.78	2.70	2.62	2.53	2.48	2.41	2.37	2.32	2.28	2.26	
24	4.26	3.40	3.01	2.78	2.62	2.51	2.43	2.36	2.30	2.26	2.22	2.18	2.13	2.09	2.02	1.98	1.94	1.89	1.86	1.82	1.80	1.76	1.74	1.73	
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.25	3.17	3.09	3.03	2.93	2.85	2.74	2.66	2.58	2.49	2.44	2.36	2.33	2.27	2.23	2.21	
25	4.24	3.38	2.99	2.76	2.60	2.49	2.41	2.34	2.28	2.24	2.20	2.16	2.11	2.06	2.00	1.96	1.92	1.87	1.84	1.80	1.77	1.74	1.72	1.71	
25	7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.32	3.21	3.13	3.05	2.99	2.89	2.81	2.70	2.62	2.54	2.45	2.40	2.32	2.29	2.23	2.19	2.17	
26	4.22	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.23	2.18	2.15	2.10	2.05	1.99	1.95	1.90	1.85	1.82	1.78	1.76	1.72	1.70	1.69	
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.17	3.09	3.02	2.96	2.86	2.77	2.66	2.58	2.50	2.41	2.36	2.28	2.25	2.19	2.15	2.13	

The function, $F = e$ with exponent 2x, is computed in part from Fisher's table VI (7). Additional entries are by interpolation, mostly graphical.

TABLE 10.7-5% (ROMAN TYPE) AND 1% (BOLD FACE TYPE) POINTS FOR THE DISTRIBUTION OF F

F ₁	n ₂ degrees of freedom (for greater mean square)																						F ₂	
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	100	200	500		∞
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.30	2.25	2.20	2.16	2.13	2.08	2.03	1.97	1.93	1.88	1.84	1.80	1.76	1.74	1.71	1.68	1.67
	7.68	5.49	4.60	4.11	3.79	3.56	3.39	3.26	3.14	3.06	2.98	2.93	2.83	2.74	2.63	2.55	2.47	2.38	2.33	2.25	2.21	2.16	2.12	2.10
28	4.20	3.34	2.95	2.71	2.56	2.44	2.36	2.29	2.24	2.19	2.15	2.12	2.06	2.02	1.96	1.91	1.87	1.81	1.78	1.75	1.72	1.69	1.67	1.65
	7.64	5.45	4.57	4.07	3.76	3.53	3.36	3.23	3.11	3.03	2.95	2.90	2.80	2.71	2.60	2.52	2.44	2.35	2.30	2.22	2.18	2.13	2.09	2.06
29	4.18	3.33	2.93	2.70	2.54	2.43	2.35	2.28	2.22	2.18	2.14	2.10	2.05	2.00	1.94	1.90	1.85	1.80	1.77	1.73	1.71	1.68	1.65	1.64
	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.08	3.00	2.92	2.87	2.77	2.68	2.57	2.49	2.41	2.32	2.27	2.19	2.15	2.10	2.06	2.03
30	4.17	3.32	2.92	2.69	2.53	2.42	2.34	2.27	2.21	2.16	2.12	2.09	2.04	1.99	1.93	1.89	1.84	1.79	1.76	1.72	1.69	1.66	1.64	1.62
	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.06	2.98	2.90	2.84	2.74	2.66	2.55	2.47	2.38	2.29	2.24	2.16	2.13	2.07	2.03	2.01
32	4.15	3.30	2.90	2.67	2.51	2.40	2.32	2.25	2.19	2.14	2.10	2.07	2.02	1.97	1.91	1.86	1.82	1.76	1.74	1.69	1.67	1.64	1.61	1.59
	7.50	5.34	4.46	3.97	3.66	3.42	3.25	3.12	3.01	2.94	2.86	2.80	2.70	2.62	2.51	2.42	2.34	2.25	2.20	2.12	2.08	2.02	1.98	1.96
34	4.13	3.28	2.88	2.65	2.49	2.38	2.30	2.23	2.17	2.12	2.08	2.05	2.00	1.95	1.89	1.84	1.80	1.74	1.71	1.67	1.64	1.61	1.59	1.57
	7.44	5.29	4.42	3.93	3.61	3.38	3.21	3.08	2.97	2.89	2.82	2.76	2.66	2.58	2.47	2.38	2.30	2.21	2.15	2.08	2.04	1.98	1.94	1.91
36	4.11	3.26	2.86	2.63	2.48	2.36	2.28	2.21	2.15	2.10	2.06	2.03	1.98	1.93	1.87	1.82	1.78	1.72	1.69	1.65	1.62	1.59	1.56	1.55
	7.39	5.25	4.38	3.89	3.58	3.35	3.18	3.04	2.94	2.86	2.78	2.72	2.62	2.54	2.43	2.35	2.26	2.17	2.12	2.04	2.00	1.94	1.90	1.87
38	4.10	3.25	2.85	2.62	2.46	2.35	2.26	2.19	2.14	2.09	2.05	2.02	1.96	1.92	1.85	1.80	1.76	1.71	1.67	1.63	1.60	1.57	1.54	1.53
	7.35	5.21	4.34	3.86	3.54	3.32	3.15	3.02	2.91	2.82	2.75	2.69	2.59	2.51	2.40	2.32	2.22	2.14	2.08	2.00	1.97	1.90	1.86	1.84
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.07	2.04	2.00	1.95	1.90	1.84	1.79	1.74	1.69	1.66	1.61	1.59	1.55	1.53	1.51
	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.88	2.80	2.73	2.66	2.56	2.49	2.37	2.29	2.20	2.11	2.05	1.97	1.94	1.88	1.84	1.81
42	4.07	3.22	2.83	2.59	2.44	2.32	2.24	2.17	2.11	2.06	2.02	1.99	1.94	1.89	1.82	1.78	1.73	1.68	1.64	1.60	1.57	1.54	1.51	1.49
	7.27	5.15	4.29	3.80	3.49	3.26	3.10	2.96	2.86	2.77	2.70	2.64	2.54	2.46	2.35	2.26	2.17	2.08	2.02	1.94	1.91	1.85	1.80	1.78
44	4.06	3.21	2.82	2.58	2.43	2.31	2.23	2.16	2.10	2.05	2.01	1.98	1.92	1.88	1.81	1.76	1.72	1.66	1.63	1.58	1.56	1.52	1.50	1.48
	7.24	5.12	4.26	3.78	3.46	3.24	3.07	2.94	2.84	2.75	2.68	2.62	2.52	2.44	2.32	2.24	2.15	2.06	2.00	1.92	1.88	1.82	1.78	1.75
46	4.05	3.20	2.81	2.57	2.42	2.30	2.22	2.14	2.09	2.04	2.00	1.97	1.91	1.87	1.80	1.75	1.71	1.65	1.62	1.57	1.54	1.51	1.48	1.46
	7.21	5.10	4.24	3.76	3.44	3.22	3.05	2.92	2.82	2.73	2.66	2.60	2.50	2.42	2.30	2.22	2.13	2.04	1.98	1.90	1.86	1.80	1.76	1.72
48	4.04	3.19	2.80	2.56	2.41	2.30	2.21	2.14	2.08	2.03	1.99	1.96	1.90	1.86	1.79	1.74	1.70	1.64	1.61	1.56	1.53	1.50	1.47	1.45
	7.19	5.08	4.22	3.74	3.42	3.20	3.04	2.90	2.80	2.71	2.64	2.58	2.48	2.40	2.28	2.20	2.11	2.02	1.96	1.88	1.84	1.78	1.73	1.70

TABLE 10.7-5% (ROMAN TYPE) AND 1% (BOLD FACE TYPE) POINTS FOR THE DISTRIBUTION OF F

F ₁	n ₂ degrees of freedom (for greater mean square)																			F ₂
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	
50	4.03 7.17	3.18 5.06	2.70 4.20	2.56 3.72	2.40 3.41	2.29 3.18	2.20 3.02	2.13 2.88	2.07 2.78	1.98 2.70	1.95 2.62	1.90 2.56	1.85 2.46	1.78 2.39	1.74 2.26	1.69 2.18	1.63 2.10	1.56 2.00	1.55 1.94	1.44 1.71
55	4.02 7.12	3.17 5.01	2.78 4.16	2.54 3.68	2.38 3.37	2.27 3.15	2.18 2.98	2.11 2.85	2.05 2.75	1.97 2.66	1.93 2.59	1.88 2.53	1.83 2.43	1.76 2.35	1.72 2.23	1.67 2.15	1.61 2.06	1.58 1.96	1.52 1.90	1.41 1.66
60	4.00 7.08	3.15 4.98	2.76 4.13	2.52 3.65	2.37 3.34	2.25 3.12	2.17 2.95	2.10 2.82	2.04 2.72	1.99 2.63	1.95 2.56	1.92 2.50	1.86 2.40	1.81 2.32	1.75 2.20	1.70 2.12	1.65 2.03	1.59 1.93	1.50 1.79	1.41 1.63
65	3.99 7.04	3.14 4.95	2.75 4.10	2.51 3.62	2.36 3.31	2.24 3.09	2.15 2.93	2.08 2.79	2.02 2.70	1.98 2.61	1.94 2.54	1.90 2.47	1.85 2.37	1.80 2.30	1.73 2.18	1.68 2.09	1.63 2.00	1.57 1.90	1.54 1.84	1.39 1.60
70	3.98 7.01	3.13 4.92	2.74 4.08	2.50 3.60	2.35 3.29	2.23 3.07	2.14 2.91	2.07 2.77	2.01 2.67	1.97 2.59	1.93 2.51	1.89 2.45	1.84 2.35	1.79 2.28	1.72 2.15	1.67 2.07	1.62 1.98	1.56 1.88	1.53 1.82	1.37 1.56
80	3.96 6.96	3.11 4.88	2.72 4.04	2.48 3.56	2.33 3.25	2.21 3.04	2.12 2.87	2.05 2.74	1.99 2.64	1.95 2.55	1.91 2.48	1.88 2.41	1.82 2.32	1.77 2.24	1.70 2.11	1.65 2.03	1.60 1.94	1.54 1.84	1.51 1.78	1.32 1.49
100	3.94 6.90	3.09 4.82	2.70 3.98	2.46 3.51	2.30 3.20	2.19 2.99	2.10 2.82	2.03 2.69	1.97 2.59	1.92 2.51	1.88 2.43	1.85 2.36	1.79 2.26	1.75 2.19	1.68 2.06	1.63 1.98	1.57 1.89	1.51 1.79	1.48 1.73	1.28 1.43
125	3.92 6.84	3.07 4.78	2.68 3.94	2.44 3.47	2.29 3.17	2.17 2.95	2.08 2.79	2.01 2.65	1.95 2.56	1.90 2.47	1.86 2.40	1.83 2.33	1.77 2.23	1.72 2.15	1.65 2.03	1.60 1.94	1.55 1.85	1.49 1.75	1.45 1.68	1.27 1.40
150	3.91 6.81	3.06 4.75	2.67 3.91	2.43 3.44	2.27 3.14	2.16 2.92	2.07 2.76	2.00 2.62	1.94 2.53	1.89 2.44	1.85 2.37	1.82 2.30	1.76 2.20	1.71 2.12	1.64 2.00	1.59 1.91	1.54 1.83	1.47 1.72	1.44 1.66	1.25 1.37
200	3.89 6.76	3.04 4.71	2.65 3.88	2.41 3.41	2.26 3.11	2.14 2.90	2.05 2.73	1.98 2.60	1.92 2.50	1.87 2.41	1.83 2.34	1.80 2.28	1.74 2.17	1.69 2.09	1.62 1.97	1.57 1.88	1.52 1.79	1.45 1.69	1.42 1.62	1.19 1.33
400	3.86 6.70	3.02 4.66	2.62 3.83	2.39 3.36	2.23 3.06	2.12 2.85	2.03 2.69	1.96 2.55	1.90 2.46	1.85 2.37	1.81 2.29	1.78 2.23	1.72 2.12	1.67 2.04	1.60 1.92	1.54 1.84	1.49 1.74	1.42 1.64	1.38 1.57	1.16 1.24
1000	3.85 6.66	3.00 4.62	2.61 3.80	2.38 3.34	2.22 3.04	2.10 2.82	2.02 2.66	1.95 2.53	1.89 2.43	1.84 2.34	1.80 2.26	1.76 2.20	1.70 2.09	1.65 2.01	1.58 1.89	1.53 1.81	1.47 1.71	1.41 1.61	1.36 1.54	1.13 1.19
∞	3.84 6.64	2.99 4.60	2.60 3.75	2.37 3.32	2.21 3.02	2.09 2.80	2.01 2.64	1.94 2.51	1.88 2.41	1.83 2.32	1.79 2.24	1.75 2.18	1.69 2.07	1.64 2.00	1.57 1.87	1.52 1.79	1.46 1.69	1.40 1.59	1.35 1.52	1.10 1.15

moderately skewed distributions. (ii) Precautions about conduct of the experiment are the same as those in section 4.3. Some pertinent evidence will be presented in section 10.9. (iii) It is usual to denote a probability between the 5% and 1% levels by an asterisk; beyond the 1% level, by a double asterisk. See tables 10.13 and 10.17.

EXAMPLE 10.4—Here are four groups of data for easy practice:

Group 1: 6, 5, 5, 2, 1, 5, 4, 2, 4, 6
 Group 2: 7, 9, 6, 8, 4, 7, 11, 8, 7, 3
 Group 3: 6, 12, 9, 13, 10, 11, 10, 8, 11, 10
 Group 4: 7, 4, 5, 3, 2, 0, 9, 7, 6, 7

Test the significance of the differences among the group means. Ans. $F = 14$.

EXAMPLE 10.5—The percentage of clean wool in 7 bags was estimated by taking 3 batches at random from each bag. The percentages of clean wool in the batches were as follows:

Batch	Bag Number						
	1	2	3	4	5	6	7
1	41.8	33.0	38.5	43.7	34.2	32.6	36.2
2	38.9	37.5	35.9	38.9	38.6	38.4	33.4
3	36.1	33.1	33.9	36.3	40.2	34.8	37.9

Show that $F = 1.35$.

EXAMPLE 10.6—In the foregoing wool data, are you willing to accept the hypothesis that the 21 batches are randomly drawn from a single, normally distributed population? If so, the mean of the entire sample is 36.85% with variance (20 *d.f.*), 9.086, figures easily verified from your calculations. Assuming all bags of the same weight, set 95% fiducial limits to the mean percentage of clean wool in the 7 bags. Ans. 35.48%–38.22%. (See example 10.12.)

EXAMPLE 10.7—In table 10.4, p stands for any group, being successively 1, 2, . . . n . A good exercise in algebra is proving the formula,

$$S(X - \bar{x})^2 = kS(\bar{x}_p - \bar{x})^2 + SS(X_p - \bar{x}_p)^2,$$

which is the addition theorem for sums of squares in analysis of variance. The left member of the equation is the sum of squares of deviations of all items from the general mean, $\bar{x}$. The first term in the right member is the sum of squares among class means, $\bar{x}_p$, representing successively the n such means. The last term is the sum of the n sums of squares of deviations of class items from the class mean. Whether you derive the formula or not, it should be carefully studied.

10.5—Analysis of variance in 2 subsamples. Chapter 4 is a special case of the present one, so it is informative to analyze the variance of the two lots of chicks studied in that chapter. As has been noted, the pooling of the subsample sums of squares is a common feature of the two chapters. For the chick data of tables 4.1 and 4.3, this pooled sum of squares, 16,220, with *d.f.* = 20 is entered in table 10.8. We shall complete the analysis of variance by computing the sum of squares for the two

TABLE 10.8

ANALYSIS OF VARIANCE OF CHICK COMB EXPERIMENT, TABLE 4.1. TWO SUBSAMPLES

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Chicks	20	16,220	811
Lot means	1	9,245.5	9,245.5

$$F = 9,245.5/811 = 11.4. \qquad \sqrt{F} = 3.38 = t$$

subsample means, 97 and 56 mg. It is perhaps clearest to calculate first the two sums, $97 \times 11 = 1,067$ and $56 \times 11 = 616$. Then, the sum of squares is, as usual,

$$\frac{(1,067)^2}{11} - \frac{(1,683)^2}{22} = 9,245.5$$

This is entered in the table and the analysis then completed. F is greater than its 1% point, 8.10, just as t in table 4.3 was beyond its.

Here are two points of interest. (i) If there is only a single degree of freedom for subsample means, then $\sqrt{F} = t$. The probability of F greater than 11.4 is identical with that of t greater than 3.38, so it is a matter of choice which one uses. (ii) With only two subsamples, the sum of squares for their means is easily calculated this way:

$$\frac{(1,067 - 616)^2}{(11)(2)} = 9,245.5,$$

as before. Merely square the difference of the two subsample sums, then divide by the total sample size. Equally, the result may be got thus:

$$\frac{11(97 - 56)^2}{2} = 9,245.5$$

I think you will agree that the F test is easier to perform than the t .

The reader interested in a short course may now go to section 10.8 and thence to chapter 11.

EXAMPLE 10.8—Hansberry and Richardson (6) gave the percentages of wormy apples on two groups of 12 trees each. The trees of group A, sprayed with lead arsenate, had 19, 26, 22, 13, 26, 25, 38, 40, 36, 12, 16, and 8 percent of apples wormy. Those of group B, sprayed with calcium arsenate and buffer materials, had 36, 42, 20, 43, 47, 49, 59, 37, 28, 49, 31, and 39 percent wormy. Compute the mean square *within groups*, 111.41, with 22 *df.*; and that *between group means*, 1650.04, with 1 *df.* Then

$$F = 1,650.04/111.41 = 14.8$$

Now, test the significance of the difference between the group means in the manner of table 4.3. The value of t is $3.85 = \sqrt{14.8}$.

10.6—Estimates of variance. The striking difference between the pig experiment of table 10.2 and the doughnuts of 10.3 is this: in the first, the mean squares for lot means and for individuals both estimate the same variance, σ^2 ; but in the second, the fat means exhibit a significantly larger variance than that of the mixes. Something has been added. This suggests a different attitude toward the two experiments. In the first, the means, all drawn from the same population, had only the variability due to random sampling from the common source; whereas, the fat means had an additional component attributed to differences among population means.

It is easy to estimate this added component, σ_m^2 , if it is assumed to be uncorrelated with σ^2 , because in that case the variance of group means is simply the sum of two variances. One is σ_m^2 , that of the fat population means; while the other is σ^2/k , the variance of the mean of the subsample containing k mixes. Thus,

$$\text{Variance of subsample means} = \sigma_m^2 + \frac{\sigma^2}{k}$$

Now σ^2 , assumed to be the same in all the populations, is estimated by the mean square for individuals, $s^2 = 141.6$ in the doughnut experiment. (As usual, s represents the sample estimate of the population σ .) Furthermore, if you remember (section 10.2) that the variance of the subsample means is (mean square)/ k , then we may write for the doughnuts,

$$\frac{503.9}{6} = s_m^2 + \frac{141.6}{6}$$

Solving this equation,

$$s_m^2 = \frac{503.9 - 141.6}{6} = 60.4$$

Thus, for estimating σ_m^2 , we have the easily remembered rule: *Subtract the mean square for individuals from the mean square for subsamples, then divide by the number of individuals per subsample.*

Don't get the idea that you can isolate σ_m^2 experimentally. The only way to evaluate a mean is through the sum of the individual measurements, and individual variation sticks to these measurements inseparably.

It is not hard to illustrate these concepts by sampling. One population can be set up to represent the subsample means with variance σ_m^2 and another to represent individuals with variance σ^2 ; then samples can be drawn from each and combined in any desired proportion. In table 10.9 is one such drawing. The gains representing the lot components are from table 3.11 in which $\sigma_m^2 = 25$: there is one of these for each lot. Individual pig gains are taken from table 3.1 where $\sigma^2 = 100$. The sum represents the experimental gain in weight set down in column 4 and analyzed in the lower part of the table.

TABLE 10.9

GAINS IN WEIGHT OF 20 PIGS IN 10 LOTS OF 2 PIGS EACH

Each gain is the sum of two components. The component for lots is a sample from table 3.11 with $\sigma^2 = 25$, while that for individuals is from table 3.1 with $\sigma^2 = 100$

Lot Number	Lot (subsample) Component	Individual Component	Pig Gains	Lot Gains
(1)	(2)	(3)	(4) = (2) + (3)	(5)
1	29	37 39	66 68	134
2	32	26 7	58 39	97
3	29	30 49	59 78	137
4	30	32 32	62 62	124
5	26	33 42	59 68	127
6	20	39 33	59 53	112
7	40	35 26	75 66	141
8	32	11 20	43 52	95
9	34	26 48	60 82	142
10	28	45 24	73 52	125

Analysis of Variance

Source of Variation	Degrees of Freedom	Mean Square
Lot means	9	144.6
Individuals	10	96.5

$$s^2 = 96.5$$

$$s_m^2 = (144.6 - 96.5)/2 = 24.0$$

In the last line of the table is the verification sought. The individual mean square is an estimate of $\sigma^2 = 100$. The variance of lot means (the lot mean square divided by k), that is, $144.6/2 = 72.3$, is an estimate of

$$\sigma_m^2 + \sigma^2/k = 25 + 100/2 = 75$$

Finally, $\sigma_m^2 = 25$ is estimated by

$$\frac{144.6 - 96.5}{2} = 24.0$$

In this way, from the analysis of variance we have recovered estimates of the two components that we put into the pig gains. There is here some sampling variation, of course, but not as much as usually may be expected: this example was chosen because of its accurate estimates. An idea of the ordinary variation can be got from examination of the records of 25 similar samples in table 10.10.

TABLE 10.10
ESTIMATES OF $\sigma_m^2 = 25$ AND $\sigma^2 = 100$ MADE FROM 25 SAMPLES DRAWN LIKE
THAT OF TABLE 10.9

Sample Number	Estimate of $\sigma_m^2 = 25$	Estimate of $\sigma^2 = 100$	Sample Number	Estimate of $\sigma_m^2 = 25$	Estimate of $\sigma^2 = 100$
1	60	127	14	56	112
2	56	104	15	-33	159
3	28	97	16	67	54
4	6	91	17	-18	90
5	18	60	18	33	65
6	-5	91	19	-21	127
7	7	53	20	-48	126
8	-1	87	21	4	43
9	0	66	22	3	145
10	-78	210	23	49	142
11	14	148	24	75	23
12	7	162	25	77	106
13	68	76	Mean	17.0	102.6

One is struck immediately by the great variability in the estimates of σ_m^2 , some of them even being negative. These latter merely indicate that the mean square for lots is less than that for individuals: the lots vary less than random samples ordinarily do if drawn from a single, normal population. Clearly, one cannot hope for accurate estimates of σ^2 and σ_m^2 from such small subsamples.

EXAMPLE 10.9—Among the 25 samples reported in table 10.10, one of the more divergent had this analysis of variance:

Lot means	9	67.7
Individuals	10	53.2

Show that σ_m^2 is estimated by 7.2.

EXAMPLE 10.10—Another sample resulted in this analysis:

Lot means	9	53.5
Individuals	10	210.4

If you are trained mathematically, you will probably be surprised by the estimate, $s_m^2 = -78.4$. Sampling variation has caused an exaggerated estimate of $\sigma^2 = 100$ and an underestimate of $\sigma^2 + 2\sigma_m^2 = 150$. If you set up the hypothesis that $\sigma_m^2 = 0$, you would have little evidence against it in this sample. For another viewpoint, see section 10.11.

EXAMPLE 10.11—Show that in the sampling of table 10.10, the value of F based on population parameters is 1.50. In the 25 samples, mean $F = 1.67$.

EXAMPLE 10.12—In the wool data of example 10.5, $s^2 = 8.22$ and $s^2 + 3s_m^2 = 11.11$. Calculate $s_m^2 = 0.96$. If you think that the hypothesis, $\sigma_m^2 = 0$ should be rejected despite the small value of F , and if you suspect that the bags were not all equal in weight, you will find that appropriate methods of estimation are discussed in chapter 17.

10.7—Estimates of variance and the F test. The foregoing section adds to one's understanding of the F test of significance. From what has been said it is plain that the mean square for subsamples is an estimate of

$$k\left(\frac{\sigma^2}{k} + \sigma_m^2\right) = \sigma^2 + k\sigma_m^2$$

It may be concluded, then, that in the populations sampled the variance ratio is

$$\frac{\text{subsample mean square}}{\text{individual mean square}} = \frac{\sigma^2 + k\sigma_m^2}{\sigma^2}$$

Hence, the hypothesis which was set up in section 10.4 implies that $\sigma_m^2 = 0$ and that F , therefore, is an estimate of 1. In verification, the mean of the F 's in table 10.6 was 0.96, and these F 's were samples from a population having the specifications of the present hypothesis.

Contrast with this the doughnut experiment. Using the sample estimates calculated in the foregoing section,

$$F = \frac{141.6 + 6(60.4)}{141.6} = 3.56$$

The additional thing now clear is the manner in which the large F is caused by the component of variance, $\sigma_m^2 = 60.4$, associated with the fat means. In the ordinary run of samples, rejection of the null hypothesis results from the added term, $k\sigma_m^2$, in the numerator of the variance ratio.

In table 10.11 is a summary of the estimates and of the test of significance provided by the analysis of variance.

TABLE 10.11
ESTIMATES AND TEST OF SIGNIFICANCE IN ANALYSIS OF VARIANCE. SUBCLASS
NUMBERS EQUAL

Source of Variation	Degrees of Freedom	If $\sigma_m^2 = 0$, Mean Square Is Estimate of	If $\sigma_m^2 \neq 0$ Mean Square Is Estimate of	If $\sigma_m^2 \neq 0$ Variance Is Estimate of
Subsample means	$n - 1$	σ^2	$\sigma^2 + k\sigma_m^2$	$\sigma^2/k + \sigma_m^2$
Individuals	$n(k - 1)$	σ^2	σ^2	σ^2
Total	$nk - 1$	σ^2		
F is estimate of		1	$\frac{\sigma^2 + k\sigma_m^2}{\sigma^2}$	

10.8—Subsamples with different numbers of individuals. Sometimes it is not possible to provide subsamples of equal size, and at other times, not desirable. As an example of the latter case, suppose one is investigating the butterfat content of the milk from cows of various breeds supplying the creameries of some district. He might wish to conduct the investigation so that the average butterfat yield per cow when multiplied by the total number of producing animals in the district would be an estimate of the total fat production. In that event, it would be necessary to sample the breeds by taking from each a number of cows proportional to the district population of that breed. Otherwise, the representatives of some rare, low yielding breed would inordinately depress the average. It is just as important to choose samples representative of the population examined as it is to make quite random the taking of the ultimate experimental units. The two procedures are not incompatible, both being integral parts of good sampling designs. Incidentally, the criterion of representativeness makes it imperative that the investigator confine his interpretations to the sampled population, not broadcasting them over populations in general.

Another source of unequal-sized subsamples is that illustrated in table 10.12. Much experimental material has such natural groupings. Eight sows bore these litters, so that differences among the mean birth weights may be attributed partly to sow individualities and partly to litter size.

So far as method of computation is concerned, only a single minor alteration is required to adapt it to the inequality in subsample sizes. This is in connection with the sum of squares between means of litters.

TABLE 10.12
BIRTH WEIGHTS (POUNDS) OF POLAND CHINA PIGS IN EIGHT LITTERS
Spring, 1919

Symbol	Litter								Total
	1	2	3	4	5	6	7	8	
$\bar{X}$	2.0	3.5	3.3	3.2	2.6	3.1	2.6	2.5	
	2.8	2.8	3.6	3.3	2.6	2.9	2.2	2.4	
	3.3	3.2	2.6	3.2	2.9	3.1	2.2	3.0	
	3.2	3.5	3.1	2.9	2.0	2.5	2.5	1.5	
	4.4	2.3	3.2	3.3	2.0		1.2		
	3.6	2.4	3.3	2.5	2.1		1.2		
	1.9	2.0	2.9	2.6					
	3.3	1.6	3.4	2.8					
	2.8		3.2						
	1.1		3.2						
ΣX	28.4	21.3	31.8	23.8	14.2	11.6	11.9	9.4	152.4
k	10	8	10	8	6	4	6	4	56
$\bar{x}$	2.84	2.66	3.18	2.98	2.37	2.90	1.98	2.35	
Σx^2	88.84	60.19	101.80	71.52	34.34	33.88	25.57	23.26	439.40

Since k changes from one litter to another, the eight quantities, $(SX)^2$, must each be divided by its own k in order to reduce it to a per-item basis. That is, $(28.4)^2/10 + (21.3)^2/8 + \dots + (9.4)^2/4 - (152.4)^2/56 = 7.48$ is the computational procedure which replaces the fourth step of section 10.3. The corresponding formula, replacing step 4 in table 10.4, is

$$S_1^2/k_1 + S_2^2/k_2 + \dots + S_n^2/k_n - C$$

The remaining results of calculation from table 10.10 are easily verified:

$$\text{Correction term} = (152.4)^2/56 = 414.75$$

$$\text{Total sum of squares} = 439.40 - 414.75 = 24.65$$

The total sum of squares and that between litters, entered in table 10.13, are subtracted to get the sum of squares for pigs of the same litter.

TABLE 10.13
ANALYSIS OF VARIANCE OF BIRTH WEIGHTS OF 56 PIGS IN 8 LITTERS OF VARIOUS SIZES

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Total	55	24.65	
Litter means	7	7.48	1.07**
Pigs of same litter	48	17.17	0.36

The degrees of freedom may be got by subtraction, also; or they may be thought of as the sum of those from the eight litters,

$$9 + 7 + 9 + 7 + 5 + 3 + 5 + 3 = 48$$

This way of thinking about degrees of freedom is valuable because it brings to mind the fact that the associated sum of squares, 17.17, is itself the sum of eight litter sums of squares, each of which may be independently computed.

The two mean squares, for litters, 1.07, and for pigs, 0.36, are notably different. Since $F = 1.07/0.36 = 2.97$ is just short of the 1% point, 3.04, it is not likely that the litter birth weights constitute random samples from a homogeneous population. As has been indicated, there are two sources of variability in the litter means. First, the dams may differ in their ability to nourish the fetuses, resulting in a variation of mean birth weights even when other factors are identical. Second, the size of the litter may affect the pig weights. Since the segregation of the sums of squares attributable to these two sources of variation involves an extension of the foregoing statistical methods, it will be postponed till the following sections. At present it is only necessary to note the fact that there are real litter differences to be examined and, if possible, explained. See, also, section 10.13.

The symbolical summary of computations for varying numbers of items in the groups is set out in table 10.14. One new feature appears.

TABLE 10.14
SYMBOLICAL SUMMARY OF COMPUTATIONS, WITH VARYING NUMBERS
OF ITEMS IN THE SUBSAMPLES

The Data					
Group 1	Group 2	...	Group p	...	Group n
X_1	X_2		X_p		X_n
$\vdots$	$\vdots$		$\vdots$		$\vdots$
k_1 items	k_2 items		k_p items		k_n items
S_1	S_2	...	S_p	...	S_n
The Computations					
1. The sum: $S = S_1 + S_2 + \dots + S_n$					
2. The correction term: $C = S^2/Sk$					
3. The total sum of squares: $SX^2 - C$					
4. Subsample means: $S_1^2/k_1 + \dots + S_n^2/k_n - C$					
5. Individuals: total - subsample means					
6. Average k : $k_o = \frac{1}{n-1} (Sk - \frac{Sk^2}{Sk})$					
Analysis and Estimates of Variance					
Source of Variation	Degrees of Freedom	If $\sigma_m^2 = 0$ Mean Square Is Estimate of	If $\sigma_m^2 \neq 0$ Mean Square Is Estimate of	If $\sigma_m^2 \neq 0$ Variance Is Estimate of	
Subsample means	$n - 1$	σ^2	$\sigma^2 + k_o \sigma_m^2$	$\sigma^2/k_o + \sigma_m^2$	
Individuals	$Sk - n$	σ^2	σ^2	σ^2	
Total	$Sk - 1$	σ^2			
F is an estimate of		1	$\frac{\sigma^2 + k_o \sigma_m^2}{\sigma^2}$		

Since the subsamples have varying k 's, an average number, k_o , is needed for estimating variance. This is,

$$k_o = \frac{1}{n-1} \left(Sk - \frac{Sk^2}{Sk} \right)$$

For the pig data, Sk is the total sample size, 56 animals. $Sk^2 = 10^2 + 10^2 + 8^2 + \dots + 4^2 = 432$. Then

$$k_o = \frac{1}{7} \left(56 - \frac{432}{56} \right) = 6.90$$

This average is always smaller than the mean of the k 's.

σ_m^2 is now estimated by

$$\frac{1.07 - 0.36}{6.90} = 0.103$$

EXAMPLE 10.13—Here is an easy example for practice:

Group 1: 18, 11, 16, 12, 14, 17, 18, 20

Group 2: 19

Group 3: 16, 18, 26

Group 4: 19, 15, 15, 10, 24, 17, 13, 17, 13, 16

The total sum of squares, 307.65 with 22 *df.*, is divided into 50.15 for group means, *df.* = 3, and 257.50 for individuals. $F = 1.23$.

EXAMPLE 10.14—Applying analysis of variance to the two lots of rat gains in section 4.5, show that $\sqrt{F} = t = 1.89$. What portions of table 4.5 are used in the present computations?

EXAMPLE 10.15—Jenkins (8) and Snedecor (14) compared the yields of a number of varieties of corn, each variety being represented by several inbred lines. Six varieties, with yields (bushels per acre) of their inbred lines, are as follows:

1. Four County: 7.3, 4.5, 7.4, 7.4, 5.0, 5.9, 6.4, 6.3, 5.0, 6.1, 7.9, 5.7

2. Silver King: 7.7, 5.4, 5.2, 4.0

3. Iodent: 6.9, 6.8, 7.6, 8.1, 9.4, 12.0, 15.9, 7.4, 9.0, 5.2, 9.2, 8.6

4. Lancaster: 9.6, 7.8, 9.6, 7.7, 8.2, 7.3, 11.3, 9.5, 8.8, 8.4, 6.8

5. Osterland: 4.8, 9.2, 8.5, 8.8, 7.9, 5.9, 9.2

6. Clark: 4.3, 8.4, 6.6, 4.9, 5.8, 7.6, 3.7

Calculate $F = 17.55/3.38 = 5.19$, *df.* = 5 and 47. Estimate the component of variance associated with the varieties of the preceding example. Ans. $k_0 = 8.63$, $s_m^2 = 1.64$.

EXAMPLE 10.16—Among 224 mice inoculated with 3 strains of typhoid organisms, the days of survival are summarized in the following frequency distributions. In applying the method of section 8.2 it is perhaps easier not to code the days-to-death, *X*, but to use the recorded data. The column of totals is necessary only for verification. Calculate $F = 179.9/5.78 = 31.1$, *df.* = 2 and 221.

Days to Death	Numbers of Mice Inoculated With Indicated Strain			Total
	9D	11C	DSC1	
2	6	1	3	10
3	4	3	5	12
4	9	3	5	17
5	8	6	8	22
6	3	6	19	28
7	1	14	23	38
8		11	22	33
9		4	14	18
10		6	14	20
11		2	7	9
12		3	8	11
13		1	4	5
14			1	1
Total	31	60	133	224
$\sum X$	125	442	1,037	1,604
$\sum X^2$	561	3,602	8,961	13,124

EXAMPLE 10.17—If k_p represents the number of items in the p th class, X_p the observed values in that class, and $\bar{x}$, the mean, then the addition theorem for sums of squares is given by

$$S(X - \bar{x})^2 = Sk_p(\bar{x}_p - \bar{x})^2 + SS(X_p - \bar{x}_p)^2$$

Since the values of k_p are not the same for all classes, each $(\bar{x}_p - \bar{x})^2$ must be multiplied by its own k_p . Contrast this formula with that of example 10.7. Can you prove the theorem? You have already verified it several times by arithmetic processes.

10.9—Analysis of variance of subsample means. Equal subsample numbers. Repeatedly, attention has been called to the fact that when lots of individuals are kept in separate compartments, then if different treatments are applied to the lots there is no way to learn whether resulting differences are due to the treatments or to other causes. The experiment summarized in table 10.15 illustrates one of the many designs that furnish information about any extraneous environmental differences that may affect the subsamples. Here, each treatment is applied to the hens

TABLE 10.15
EGGS LAID BY EACH OF 120 HENS. SIX TREATMENTS, EACH APPLIED TO 2 LOTS
OF 10 HENS. EACH LOT IN SEPARATE CAGE

Treatment	1		2		3		4		5		6	
Lot	1	2	1	2	1	2	1	2	1	2	1	2
1	100	70	94	90	84	88	82	78	87	87	70	64
2	51	90	86	88	52	86	72	86	66	79	55	56
3	39	102	52	78	103	77	72	93	50	41	71	26
4	56	81	81	102	61	66	77	87	92	82	74	58
5	56	83	60	45	63	91	67	63	96	62	64	61
6	77	98	84	90	84	45	61	72	67	83	51	59
7	75	85	81	58	78	83	29	50	85	100	42	53
8	73	90	97	70	51	106	36	81	80	51	86	72
9	65	90	81	65	76	88	58	78	76	85	79	48
10	78	22	78	114	32	63	64	60	86	78	68	25
Total	670	811	794	800	684	793	618	748	785	748	660	522
Lot means	67.0	81.1	79.4	80.0	68.4	79.3	61.8	74.8	78.5	74.8	66.0	52.2
Treatment Sums	1,481		1,594		1,477		1,366		1,533		1,182	
Means	74.0		79.7		73.8		68.3		76.6		59.1	

Preliminary Analysis of Variance

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Total	119	40,992.6	
Cages	11	8,827.9	
Individuals	108	32,164.7	297.8

TABLE 10.15 — (Continued)

Analysis of variance of Cage Means

Cages	11	8,827.9	
Treatments	5	5,372.3	1,074.5
Cages treated alike	6	3,455.6	575.9

Completed Analysis of Variance

Treatment	5	5,372.3	1,074.5
Cages treated alike	6	3,455.6	575.9
Individuals	108	32,164.7	297.8
Total	119	40,992.6	

in two cages. If these identically treated cages are randomly chosen, any differences among the numbers of eggs laid in them, exceeding normal sampling variation, are attributable to differences in environment, such as light, heat and air circulation. In the present experiment, these differences, if any, are included in the experimental error. Later we shall explain more efficient designs that will tend to exclude them.

In the preliminary analysis, the total sum of squares is separated, in the usual fashion, into parts due to cage and individual variation. But it is clear that cage differences should be distinguished. One source is treatments, another is those unexplained causes that may produce differences in egg production (in excess of random fluctuation) between cages treated alike. The sum of squares for the former is got from the treatment sums by the familiar procedure,

$$\frac{(1,481)^2 + (1,594)^2 + \dots + (1,182)^2}{20} - \frac{(8,633)^2}{120} = 5,372.3,$$

the correction for mean being the same as that used in the preliminary analysis. This result leads to the analysis of variance of cage means in the table, and thence to the completed analysis.

Note on computation: The method just outlined is the customary one, the sum of squares for the overall classification, cages, being calculated first; that for the major classification, treatments, second; and finally the remainder, corresponding to cages treated alike. But the latter sum of squares can be calculated independently. Since there are only two subsamples per treatment, the method of section 10.5 is applicable. Thus, for the first treatment, the single degree of freedom for cage means has the sum of squares (or mean square),

$$\frac{(811 - 670)^2}{2(10)} = 994.05$$

The other 5 *d.f.* have the sums of squares: 1.80, 594.05, 845.00, 68.45, and 952.20, the sum being 3,455.55, rounded to 3,455.6 in the table. This computation has three merits: it verifies the tabular results, it reflects the striking contrasts among the differences between means of cages receiving the same treatment, and it emphasizes the theoretical fact that each degree of freedom has a sum of squares all its own. Heterogeneity of error variance will be discussed in section 10.13, while the third point is the subject of chapter 15.

The completed analysis of variance increases the suspicion that different environments of the cages treated alike have introduced a component of variance into the mean square for cages. In addition another component may have been contributed by treatments. As a guide to

TABLE 10.16
ESTIMATES OF COMPONENTS OF VARIANCE WITH MAJOR AND MINOR CLASSIFICATIONS
OF SUBSAMPLE MEANS. EGG LAYING DATA

Source of Variation	Mean Square	Individuals per Subclass	Mean Square Is an Estimate of	Variance Is an Estimate of
Treatment	1,074.5	20	$\sigma^2 + 10\sigma_c^2 + 20\sigma_T^2$	$\sigma^2/20 + \sigma_c^2/2 + \sigma_T^2$
Cage	575.9	10	$\sigma^2 + 10\sigma_c^2$	$\sigma^2/10 + \sigma_c^2$
Individual	297.8	1	σ^2	σ^2

$$s^2 = 297.8, \quad s_c^2 = (575.9 - 297.8)/10 = 27.8, \quad s_T^2 = (1,074.5 - 575.9)/20 = 24.9$$

appropriate tests of significance, it is convenient to extend the estimates of variance (section 10.6) to include the newly made analysis of subsample means. This is done in table 10.16. The new feature is the addition of a third component of variance by the major classification, treatment. All parts of the table will be clear upon inspection.

It is now evident that the cage component of variance, $s_c^2 = 27.8$, is almost as great as the variance of the cage mean, $s^2/10 = 29.78$. Clearly, the variation of the individual birds is only part of the experimental error. Apparently the hens were influenced in their production of eggs not only by the treatments imposed but also by some extraneous conditions in the cages. This means that even in a uniformity trial with no real treatments there would be two sources of variation in mean cage production, the normal variation of the birds and the environments peculiar to the several cages. Both these sources enter the variance attributable to treatments and experimental error as well.

There are two null hypotheses whose testing is indicated in table 10.16. The one of prime interest is $\sigma_T^2 = 0$, tested by the ratio,

$$\frac{\sigma^2 + 10\sigma_c^2 + 20\sigma_T^2}{\sigma^2 + 10\sigma_c^2}$$

This is estimated by $F = 1,074.5/575.9 = 1.87$, with $d.f. = 5$ and 6 . Since $F_{.05} = 4.39$, there is little evidence of treatment effects.

The experimenter probably would be interested in testing also the hypothesis, $\sigma_c^2 = 0$. For this, the variance ratio,

$$\frac{\sigma^2 + 10\sigma_c^2}{\sigma^2},$$

is estimated by $F = 575.9/297.8 = 1.93$, $d.f. = 6$, and 108 , $F_{.05} = 2.19$. He might reason this way: "While F is not significant, I have noticed that environmental effects usually have appeared in the past. I think σ_c^2 really exists, and I shall re-design both the battery and my experiment in order to avoid such effects in the future." Or, his reaction might be: "The laying battery has been used successfully many times, and this is the first indication of σ_c^2 that has been noticed. I think it is merely a freak of sampling and that σ_c^2 is really zero. (Remember the sampling variation of table 10.10.) Of course, I know that something unusual may have occurred this time, like an undetected disease in several of the cages, and that the experimental treatments may not have caused the differences, but nevertheless I am willing to accept the hypothesis."

If the investigator's experience and good judgment dictate this hypothesis, then table 10.16 may be revised. Replacing σ_c^2 by zero, you will see that the sums of squares for cages and individuals are estimates of the same variance, σ^2 , and may be combined into a more reliable estimate. This is done by pooling the degrees of freedom (table 10.15), $6 + 108 = 114$, and the sums of squares, $3,455.6 + 32,164.7 = 35,620.3$, then calculating the mean square, $35,620.3/114 = 312.5$. The new analysis is in table 10.17. Thus if the experimenter's decision is correct, treatment

TABLE 10.17
ANALYSIS OF VARIANCE OF EGG LAYING DATA UNDER THE HYPOTHESIS, $\sigma_c^2 = 0$

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Treatment	5	5,372.3	1,074.5
Individuals	114	35,620.3	312.5

$$F = 1,074.5/312.5 = 3.44^{**}$$

effects are significant. To see clearly one thing that is implied, observe that there are now $6 \times 19 = 114$ $d.f.$ for individuals: the 20 birds in each pair of cages treated alike are considered as a single lot. If there is any extra environmental difference between the pairs of cages it has been included in the estimate of σ^2 , increasing it from 297.8 to 312.5. But there is a second implication: the effects of any extraneous differences among

the cages treated differently are now considered as part of the treatment effects. I suspect that, under these circumstances, most investigators would hesitate in making decisions about treatments based on the auxiliary hypothesis, $\sigma_c^2 = 0$. Instead, they would probably set about the reconstruction of the laying battery, adapting it to a more effective experimental design.

What is this more effective design? Not one but several will be presented later. At the student's present stage of advancement he will appreciate one suggestion: use more cages with fewer birds per cage. This will allow more replications of each treatment. Further details will be found in example 10.26.

10.10—Analysis of variance of subsample means. Unequal subsample numbers. It was by design that the cages of laying hens could be grouped according to treatment, but the pig litters of section 10.8 have a natural grouping by size. The 7 *d.f.* for litter means may be divided into 3 corresponding to size of litter, and 4 between litters of the same size. The only change in calculation is the usual one for unequal subclass sizes: division of the squared sum is by the corresponding *k*. Since the sums of the pig weights in the 4 litter-size groups, easily computed from table 10.12, are 60.2, 45.1, 26.1, and 21.0 pounds, the sum of squares is $(60.2)^2/20 + (45.1)^2/16 + (26.1)^2/12 + (21.0)^2/8 - (152.4)^2/56 = 5.47$. The remainder, $7.48 - 5.47 = 2.01$, is attributable to the variation of mean birth weight between pairs of litters of the same size. The completed analysis is in table 10.18.

TABLE 10.18
COMPLETED ANALYSIS OF VARIANCE OF THE PIG DATA OF TABLES 10.12 AND 10.13

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Size of litters	3	5.47	1.82
Litters of same size	4	2.01	0.50
Individuals	48	17.17	0.36

Estimates of Variance With Major and Minor Classification
of Subsample Means

Source of Variation	Mean Square	Individuals per Subclass	Mean Square Is an Estimate of	Variance Is an Estimate of
Litter sizes	1.82	13.52	$\sigma^2 + 6.90\sigma_L^2 + 13.52\sigma_B^2$	$\sigma^2/13.52 + \sigma_L^2/1.96 + \sigma_B^2$
Litters	0.50	6.90	$\sigma^2 + 6.90\sigma_L^2$	$\sigma^2/6.90 + \sigma_L^2$
Individuals	0.36	1	σ^2	σ^2

$$s^2 = 0.36, \quad s_L^2 = 0.020, \quad s_B^2 = 0.098$$

For the estimates of variance in table 10.18 two averages, k_o , are required. The first, for litters, is the same as that calculated in section 10.8, $k_o = 6.90$. The second, for the size classes, 10, 8, 6, and 4 pigs, has $n = 4$, $Sk = 20 + 16 + 12 + 8 = 56$ and $Sk^2 = (20)^2 + (16)^2 + (12)^2 + (8)^2 = 864$. Whence,

$$k_o = \frac{1}{4 - 1} \left(56 - \frac{864}{56} \right) = 13.52$$

The variation between litters of the same size derives from two sources, differences among the dams producing the litters and the normal sampling variation associated with the pig variance, 0.36. The hypothesis that the population variance of the former, σ_L^2 , is zero is tested by

$$F = 0.50/0.36 = 1.4, \quad d.f. = 4 \text{ and } 48, \quad F_{.05} = 2.56$$

So far as the sample is competent, there is little evidence of differential ability to produce pigs of distinctive birth weights in litters of the same size.

Variation among pig birth weights in litters of differing sizes involves not only both sources evaluated above, but also possible differences in available nutrients during the pre-natal period. It might be supposed that total birth weight would tend to be constant, with the result that pigs in small litters would be heavier than those in large. The hypothesis, $\sigma_s^2 = 0$, is tested by

$$F = 1.82/0.50 = 3.64, \quad d.f. = 3 \text{ and } 4, \quad F_{.05} = 6.59$$

The evidence against this hypothesis, then, is slight; but examination of the litter-size means, easily calculated from table 10.12 as 3.01, 2.82, 2.18, and 2.62 pounds, reveals a trend toward lighter pigs in the smaller litters, and this doesn't seem reasonable. Apparently, the data include too few litters for decisions about these variable characteristics of birth weight.

EXAMPLE 10.18—In example 6.21 we made two estimates of the population variance of soybean heights. If you refer to the original table, 6.1, you can make the following analysis of variance:

Weeks	4	3,120.0	
Individuals	15	130.0	8.67

The 4 *d.f.* for weekly means may now be subdivided in this fashion:

Attributable to regression	1	3,097.6	
Deviations from regression	3	22.4	7.47

These latter sums of squares may be taken from table 6.3 if it is remembered that the variance of individuals is k times that of the means in that table. Thus, the sum of squares for regression is 4 times

$$(S_{xy})^2/Sx^2 = (88)^2/10 = 774.4,$$

while that for deviations is $4(780 - 774.4) = 22.4$. The two estimates of σ^2 , 8.67 and 7.47, are the same as those derived in example 6.21.

In the present example, one feature of the experimental design has been ignored—the arrangement of the pots in 4 blocks. This will be remedied in the next chapter.

EXAMPLE 10.19—The following data were abstracted from records of performance of Poland China swine in a single inbred line at the Iowa Agricultural Experiment Station. Four pigs were taken from each of 23 litters and fed a standard ration from weaning to about 225 pounds. In each litter there were chosen 2 boars and 2 sows. Every entry in the table is the sum of the average daily gains (pounds per day) of 2 pigs of the same sex:

Sex	1932				1933		
	Sire 1	Sire 2	Sire 3	Sire 4	Sire 5	Sire 6	Sire 7
Boars	2.768	2.550	2.284	2.357	2.869	2.743	2.878
Sows	2.375	2.574	2.223	2.710	2.460	2.557	2.622
Boars	2.620		3.009	2.716	2.309	2.496	3.055
Sows	2.617		2.609	2.742	2.237	2.483	2.800
Boars	2.563		2.776		2.959		2.777
Sows	2.231		2.476		2.893		2.706
Boars	2.575		1.964				2.918
Sows	2.939		2.294				2.674
Boars	2.807						
Sows	2.203						
Boars	2.789						
Sows	2.996						
Boars	2.591						
Sows	2.520						

The analysis of variance taken from the original data is:

Litters	22	0.987070
Pigs	69	1.109369
Total	91	2.096439

Verify the sum of squares for litters by use of litter sums calculated from the table.

EXAMPLE 10.20—Subdivide the sum of squares and degrees of freedom for litters as follows:

Years (Y)	1	0.083986
Sires within years (B)	5	0.169271
Litters with common sire (L)	16	0.733813

EXAMPLE 10.21—Subdivide the sum of squares and degrees of freedom for pigs thus:

Sex within litters (S)	23	0.426272
Pigs of same sex and litter	46	0.683097

EXAMPLE 10.22—From the foregoing analyses, estimate the components of variance associated with

Years	$k_o = 43.826$	$s_Y^2 = 0.001144$
Sires	12.464	$s_B^2 = -0.000963$
Litters	4	$s_L^2 = 0.006832$
Sex	2	$s_S^2 = 0.001842$
Pigs	1	$s^2 = 0.014850$

The mean square for sires is less than that for litters, probably as a result of sampling variation. It is reasonable to think that there are real differences in rate of gain

among litters by different sires, but this sample furnishes no estimate of σ_B^2 . So far as this sample is reliable, the best estimate of σ_B^2 is zero.

Geneticists will recognize the fact that $s^2 = 0.014850$ is the estimate of genetic variance least affected by environment.

10.11—Intraclass correlation. It has been shown that the variance in a sample may be separated into two parts, one reflecting the natural variation of individuals treated alike, and the other arising from genetic or environmental conditions peculiar to the subsamples. The former is an estimate of σ^2 , which may be assumed equal in the sampled populations, while the latter is an estimate of σ_m^2 , the added portion related to differences among the population means. The sum, $\sigma^2 + \sigma_m^2$, is the variance of individuals picked at random from the entire universe made up of the sampled populations. The ratio of the two variances,

$$\frac{\sigma_m^2}{\sigma^2 + \sigma_m^2},$$

is known as the *intraclass correlation*, ρ_I . In the numerator is the variance which, in a subsample, is common to all individuals because all of them are affected alike by the variation which pertains to the subsample as a whole. This is $\sigma_m^2 = 25$ in the illustrative sampling of table 10.9. In the denominator is an average variance that would apply to individuals if they were picked at random from the universe without attention to population boundaries.

The justification for calling this "correlation" may be understood by reference to section 7.4 where it was shown that r is the ratio of two variances. The numerator in both correlations is associated with the variance that is common to individuals: the *covariance* in one case, and σ_m^2 in the other. The denominator is an average of the variation that would affect individuals if they were not classified—into X_1 and X_2 , or into subsamples.

If all subsamples are from the same population, as in the pig gains of tables 10.1 and 10.6, then $\sigma_m^2 = 0$ and consequently $\rho_I = 0$: the entire variation, both individual and group, arises from the random sampling of a single population. Of course, in any particular sample from $\rho_I = 0$ the value of the intraclass correlation r_I , is not usually zero. Thus, in table 10.5, $s^2 = 114.0$, $s_m^2 = (172.3 - 114.0)/10 = 5.83$, and

$$r_I = \frac{5.83}{114.0 + 5.83} = 0.049$$

The hypothesis that $\rho_I = 0$ is identical with the hypothesis, $\sigma_m^2 = 0$, and

$$F = 172.3/114.0 = 1.51,$$

a nonsignificant value, is the appropriate test of both of them.

Large intraclass correlation indicates relatively small variation among

the individuals of the subsamples. Indeed, in the limiting situation, $\sigma^2 = 0$ and

$$\rho_I = \frac{\sigma_m^2}{0 + \sigma_m^2} = 1$$

Since this could happen only if all the individuals in each group had exactly the same measurement, it is of only theoretical interest. However, the data on identical twins in table 10.19 illustrate a striking approach

TABLE 10.19
NUMBER OF FINGER RIDGES ON BOTH HANDS OF INDIVIDUALS IN 12 PAIRS
OF FEMALE IDENTICAL TWINS
Data from Newman, Freeman and Holzinger (11)

Pair	Finger Ridges of Individuals	Pair	Finger Ridges of Individuals	Pair	Finger Ridges of Individuals
1	71, 71	5	76, 70	9	114, 113
2	79, 82	6	83, 82	10	94, 91
3	105, 99	7	114, 113	11	75, 83
4	115, 114	8	57, 44	12	76, 72

Analysis of Variance		
Source of Variation	Degrees of Freedom	Mean Square
Twin pairs	11	817.314
Individuals	12	14.292

$s^2 = 14.292, \quad s_m^2 = 401.51, \quad r_I = 0.966$

to perfect correlation. The totals of finger ridges are much the same for the two members of each pair but differ markedly among pairs. Since these counts are presumably unaffected by age, the high twin correlation, $r_I = 0.966$, measures the similarity of their inheritance of this character.

One can get a good idea of the variation of r_I in small samples by computing its values from the random samples of pig variance in table 10.10. In sample 1, for example,

$$r_I = 60/(60 + 127) = 0.321$$

Among the 25 samples the smallest intraclass correlation is -0.615 in number 20, the largest, 0.765 in number 24. This is ordinary sampling variation from the population value,

$$\rho_I = 25/(25 + 100) = 0.2$$

As in interclass correlation, there is a small bias in the intraclass. This may be corrected approximately by transformation to z (figure 7.4) and

addition of $1/(2n - 1)$, where n is the number of pairs. Thus, in sample 1 above,

$$\begin{array}{r} r_I = 0.321 \\ z = 0.333 \\ 1/(2n - 1) = 0.053 \\ \hline \text{Unbiased } z = 0.386 \\ \text{Unbiased } r_I = 0.368 \end{array}$$

If $k = 2$, intraclass correlations may be averaged and may have fiducial limits set as in section 7.6 (4, Chapter 7).

In order to calculate intraclass correlation directly from a table of analysis of variance, it is convenient to put the formula in this form:

$$r_I = \frac{M_{\bar{x}} - M}{M_{\bar{x}} + (k - 1)M},$$

where $M_{\bar{x}}$ and M denote mean squares for subsample means and for individuals. In the twin data, $M_{\bar{x}} = 817.314$, $M = 14.292$ and $k = 2$. Hence

$$r_I = \frac{817.314 - 14.292}{817.314 + (2 - 1)(14.292)} = 0.97 \text{ (as before)}$$

If the subsamples differ in size, the average value, k_o , is used. For example, in the pig data of tables 10.12 and 10.13,

$$r_I = \frac{1.07 - 0.36}{1.07 + (6.90 - 1)(0.36)} = 0.222$$

The F -test for table 10.13 shows that this small correlation is significant.

If the subsample means vary less than would be expected from the variation of the individuals, then $M_{\bar{x}}$ is smaller than M , σ_m^2 is estimated by a negative number and r_I is negative. But r_I cannot approach -1 unless $k = 2$. As can be seen from the formula,

$$\text{if } M_{\bar{x}} = 0, r_I = -\frac{1}{k - 1}$$

This indicates why the distribution of r_I is different from that of r so that the test of significance in table 7.3 cannot be applied to intraclass correlation.

If one attempts to graph the twin data, he will not know, in any one pair, which number of finger ridges is X_1 and which, X_2 . The solution is to locate 2 points for each pair: the second pair in table 10.19, for example, determines the points (79, 82) and (82, 79), symmetrically located with respect to a 45° line through the origin (45° , that is, if the scales for X_1 and X_2 are equal). If there were triplets instead of twins, each set of observations, a , b , and c , would specify 6 points, (a, b) , (b, a) , (a, c) , (c, a) , (b, c) , and (c, b) . The number of points per subsample rises rapidly with k ;

12 for $k = 4$, 20 for $k = 5$, etc. For large k 's, the job of computing r_I directly from all these pairs was discouraging until Harris (7) discovered a shortened process similar to the later developed analysis of variance (5).

EXAMPLE 10.23—Calculate the unbiased estimate of the correlation between numbers of finger ridges in table 10.19. Ans. 0.969.

EXAMPLE 10.24—From the pig-litter analysis in example 10.19 one can calculate $M = 0.01607$ and $M_s = 0.04487$. Bearing in mind the 4 pigs per litter, compute $r_I = 0.31$. This may be referred to as the intralitter correlation ignoring the effect of sex.

EXAMPLE 10.25—The more critical analysis of example 10.22 affords a number of intraclass correlations that may be of interest. The intrasex correlation is

$$r_I = \frac{0.001842}{0.001842 + 0.014850} = 0.110$$

This is small for weight gains in swine, but would be larger among poultry or rats. The intralitter correlation is

$$\frac{s_L^2}{s_L^2 + s_S^2 + s^2} = \frac{0.006832}{0.006832 + 0.001842 + 0.014850} = 0.290$$

Interest might be centered in the intralitter correlation among sows only, or among boars. The original data could be segregated by sex, then reanalyzed. But an average correlation could be arrived at in this way:

$$\frac{s_L^2}{s_L^2 + s^2} = \frac{0.006832}{0.006832 + 0.014850} = 0.315$$

Here, the sampling is within sex only, so that the sex component of variance is omitted from the denominator.

10.12—Efficiency of an experiment. After he completes the examination of some experimental data, the investigator may wonder if the experiment might have been conducted more efficiently. The most efficient experiment yields the maximum amount of information per dollar spent. One rather simple problem of this type is illustrated in table 10.20, the results of an effort to learn something about the variability of the concentration of carbon dioxide in soil air taken from fallow plots of Clarion loam (13). Triplicate determinations were made at each of 3 positions per plot. From the table it is clear that the chief source of variation is positional: the three determinations in the same place are much alike. The question to be considered is whether one could have got more information about plot differences by spreading the 9 soundings at random over each plot.

In this type of investigation, one is usually not concerned with the plot component of variance because it is an accident of the particular plots selected and may not be encountered again. Rather, one asks, "What is the average variance of plot means based on 9 determinations per plot?" The first answer is calculated for the experiment as it was conducted. The variance of a positional mean was $s^2/3 + s_P^2 = 48.82/3 = 16.27$; hence, the variance of the mean of 3 positions, assuming random selection, was $16.27/3 = 5.42$. Second, we may estimate the variance

of a plot mean as it might be if the determinations were randomized in the plots. The variance of such determinations was shown (section 10.11) to be estimated by $s^2 + s_P^2 = 1.52 + 15.77 = 17.29$. The corresponding variance of the plot mean is $17.29/9 = 1.92$. Remembering that the

TABLE 10.20
CONCENTRATION OF CARBON DIOXIDE IN THE SOIL AIR OF FALLOW PLOTS OF
CLARION LOAM
Parts per 10,000

Previous Treatment of Plot	Position on Plot			Plot Total
	1	2	3	
None	16	7	12	99
	14	10	8	
	14	8	10	
	—	—	—	
	44	25	30	
1,500 pounds per acre alfalfa plowed under	8	16	6	86
	6	16	6	
	6	16	6	
	—	—	—	
	20	48	18	
1,500 pounds per acre oat straw plowed under	11	8	6	78
	12	6	6	
	12	8	9	
	—	—	—	
	35	22	21	
Total				263

Analysis of Variance of Concentration of
Carbon Dioxide in Soil Air

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Total	26	345.19	
Plots	2	24.97	12.48
Positions within the plots	6	292.89	48.82**
Determinations at same positions	18	27.33	1.52

Components of Variance

Source of Variation	Mean Square	Number of Determinations	Mean Square Is an Estimate of	Variance Is an Estimate of
Positions	48.82	3	$\sigma^2 + 3\sigma_P^2$	$\sigma^2/3 + \sigma_P^2$
Determinations	1.52	1	σ^2	σ^2

$$s^2 = 1.52, s_P^2 = (48.82 - 1.52)/3 = 15.77$$

information from an experiment has been defined as inversely proportional to the variance, we reach the conclusion that the information could have been increased by

$$\frac{\frac{1}{1.92} - \frac{1}{5.42}}{\frac{1}{5.42}} = 182\%$$

if the soundings had been scattered randomly over the plots. You may object that it would cost more to randomize the determinations, and this may be true. The exact calculation of maximum information per unit cost is beyond the scope of this text, but we shall consider other features of it, and cite references, in chapter 17. At the moment we have learned that, in this instance, determinations in triplicate were statistically inefficient as compared to the same number of determinations made singly. This is because of the small variation in the measurement of concentration as compared to the positional variability of the concentration itself.

Owing to the prevalence of multiple determinations, we should perhaps emphasize the method just used. Let us consider three stages that are common. First, two or more samples of some material are collected at different places. Next, there are randomly taken for analysis two or more subsamples from each sample. Finally, two or more readings are made on each subsample. The estimates are like this:

Samples	$\sigma^2 + k_1\sigma_B^2 + k_2\sigma_S^2$
Subsamples	$\sigma^2 + k_1\sigma_B^2$
Readings	σ^2

The readings may be made on a single batch of material that has undergone some more or less elaborate chemical or biological processing: the variation among them, like that among the triplicate determinations in table 10.20, may have only minor sources such as differences in counts of bacteria on parts of the same plate. Subsamples, however, would presumably be prepared separately. They would differ not only because of the usual sampling variation in the material itself, such as from one position to another in table 10.20, but also because of small irregularities in the laboratory procedures. Together, these make up σ_B^2 . This kind of variation is illustrated by blood counts made from two different punctures of the same individual, the material from the punctures being handled, say, by two technicians. Finally, there is the variation among sources of the material. These might be truck loads of ear corn, a subsample being drawn from each load, or churnings of butter from each of several vats of cream. If this cumulative variation among samples is the source of experimental error, the effort should be made to adjust k_1 and k_2 so that maximum information may be obtained for the time and effort involved.

In table 10.20 there is another feature that deserves notice. The intraposition correlation within the plots is negative—

$$r_1 = \frac{12.48 - 48.82}{12.48 + (3 - 1)48.82} = -0.33$$

One may ask if this correlation is significant. But the value of F , $12.48/48.82 = 0.26$, is not found in the table. The test is easily made, however, by this device: compare the reciprocal of F , $48.82/12.48 = 3.91$, with the tabular values corresponding to the degrees of freedom *in reverse*; that is, $n_1 = 6$ and $n_2 = 2$. The 5% point, 19.33, suggests that our negative correlation may be merely an accident of sampling from a population in which $\rho_r = 0$. Perhaps if one extended his investigation to other plots with the same cultural history, all on Clarion loam, of course, he would still find no evidence of a plot component of variance.

EXAMPLE 10.26—What would be the effect on the variance of treatment means if the egg laying experiment were carried out with 4 cages per treatment, 5 hens per cage? Assuming (table 10.16) that $s^2 = 297.8$ and $s_e^2 = 27.8$ remain unchanged, the variance of the cage mean would increase to $s^2/5 + s_e^2 = 87.4$ instead of $575.9/10 = 57.6$. But the variance of the mean of 4 such cages would be $87.4/4 = 21.8$, a decrease of 24% from the variance of the mean of the 2 cages used in the experiment. This indicates 24% more information from the same birds, or 24% less birds for the same information.

10.13—Tests of homogeneity of variances. From time to time we have raised the question as to whether two or more variances may differ significantly. For instance, in example 4.3 the variances in two lots of honeybee data seemed notably different, and you were cited to the present section for a test. Since the lots were equal in numbers of bees, the F -test is used with only this modification: calculate F as the quotient of the larger variance by the smaller, then double the tabular probabilities. The first part of the rule arises from inability to label either variance with the sources, “subsample means” or “individuals.” The second follows from the fact that F so calculated is always greater than 1, so that only the upper part of the distribution is used. In the two lots of bees the variances were 0.0267 and 0.5889, hence

$$F = 0.5889/0.0267 = 22.06,$$

far beyond what is now the 2% point, 5.35.

If n_1 is not equal to n_2 , assign n_1 to the larger mean square. Consider the other sample of bee data in table 4.7. The two variances were 0.0653 and 0.4133 with *d.f.* = 19 and 9. F then is $0.4133/0.0653$ with $n_1 = 9$ and $n_2 = 19$, to be compared with the 2% point, 3.52. Since the sample F is larger than this tabular value, the null hypothesis may safely be rejected.

Bartlett’s test of homogeneity (1) is available if there are either 2 or more samples of variance, the tests indicated above being suitable for 2 samples only. If the samples are equal in size, the test involves the comparison of n times the logarithm of the average variance with the sum of

logarithms of the separate variances. In table 10.21 the method is applied to the 4 subsamples of s^2 in table 10.1—subsamples known to have been drawn from a common source. The factor, 2.3026, is a constant ($\log_{10} 10$) necessary because common logarithms are used. Chi-square, as first calculated, is slightly biased upwards. In this example, even the biased value shows less than average sampling variation in s^2 (table 9.1). It is necessary to make the correction only if χ^2 lies close above one of the

TABLE 10.21
COMPUTATION OF BARTLETT'S TEST OF HOMOGENEITY OF VARIANCE.
SAMPLES EQUAL IN SIZE
Data from table 10.1: $n = 4$, $k = 5$

Sample	Sum of Squares, Sx^2	Variance, s^2	$\log s^2$
1	472	118	2.07188
2	396	99	1.99564
3	616	154	2.18752
4	164	41	1.61278
		$Ss^2 = 412$	$S \log s^2 = 7.86782$
Mean variance = $\bar{s}^2 = Ss^2/n = 412/4 = 103$			$\log \bar{s}^2 = 2.01284$
		$n \log \bar{s}^2 = 4(2.01284) = 8.05136$	
		$S \log s^2 = 7.86782$	
		Difference	0.18354
		$\chi^2 = 2.3026 (k - 1)(n \log \bar{s}^2 - S \log s^2)$	
		$= 2.3026 (5 - 1)(0.18354)$	
		$= 1.69$, $d.f. = n - 1 = 3$.	
Correction factor = $C = 1 + \frac{n+1}{3n(k-1)} = 1 + \frac{4+1}{3(4)(5-1)} = 1.1042$			
Corrected $\chi^2 = \chi^2/C = 1.69/1.1042 = 1.53$			

critical tabular values, and then only if one wishes to get an accurate evaluation of P .

For samples of differing size, the computation of chi-square is more complicated though following the same pattern. In table 10.22 the test of homogeneity is applied to the pig birth weights of table 10.12. Each Sx^2 is calculated from the corresponding SX^2 by deducting the usual $(SX)^2/k$. The table requires no further explanation.

Another anomaly in the pig birth weights now stands out: the differing variances do not seem to have any physiological explanation. Both the largest and the smallest were in the 10-pig litters, and the others appear to occur at random. Coupled with the unexplained sizes of the birth weights (section 10.10), this leads to the conclusion that our sample contains little information.

In addition to these biological peculiarities, there is a statistical feature that modifies some of the assumptions made in section 10.10. The estimate

TABLE 10.22
COMPUTATION OF BARTLETT'S TEST OF HOMOGENEITY OF VARIANCE.
SAMPLES DIFFERING IN SIZE
Pig birth weights from table 10.12

Sample	Sx^2	Degrees of Freedom $k-1$	Reciprocal $1/(k-1)$	Variance s^2	$\log s^2$	$(k-1)\log s^2$
1	8.18	9	0.11111	0.9089	-0.04148	- 0.3733
2	3.48	7	0.14286	0.4972	-0.30347	- 2.1243
3	0.68	9	0.11111	0.0756	-1.12148	-10.0933
4	0.72	7	0.14286	0.1029	-0.98758	- 6.9131
5	0.73	5	0.20000	0.1460	-0.83565	- 4.1782
6	0.24	3	0.33333	0.0800	-1.09691	- 3.2907
7	1.97	5	0.20000	0.3940	-0.40450	-2.0225
8	1.17	3	0.33333	0.3900	-0.40894	- 1.2268
$n = 8$	17.17 $= Sx^2$	48 $= S(k-1)$	1.5746 $= S \frac{1}{k-1}$		$S(k-1)(\log s^2) = -30.2222$	
$s^2 = Sx^2/S(k-1) = 17.17/48 = 0.3577$ $(\log s^2)S(k-1) = (-0.44648)(48) = -21.4310$ $\chi^2 = 2.3026[(\log s^2)S(k-1) - S(k-1)(\log s^2)]$ $= 2.3026[-21.4310 - (-30.222)]$ $= 20.24$						
Correction factor, $C = 1 + \frac{1}{3(n-1)} \left[S \frac{1}{k-1} - \frac{1}{S(k-1)} \right]$ $= 1 + \frac{1}{(3)(7)} \left[1.5746 - \frac{1}{48} \right] = 1.074$ Corrected $\chi^2 = 20.24/1.074 = 18.85^{**}$ $df. = n - 1 = 7.$						

of individual variance, $s^2 = 0.36$, is an average of variances that may be drawn from different population values of σ^2 . Hence, the hypothesis tested in table 10.18 is not the simple one, $\sigma_m^2 = 0$, but the more general one—"The litters are drawn at random from the same normal population; that is, there is no population difference in either m or σ ." Since we now have evidence that the second part of the hypothesis is untrue, we can no longer be certain about the first part.

As for the pig litter data, the matter is unimportant: there was no understandable trend in either the birth weights or the variances, so that the sample appears uninformative. In other experiments there would be several considerations of interest. The first question that should be answered is, "What is the meaning of the different variances?" Pursuit of that lead might be more profitable than any findings about the means. If not, the natural conclusion would seem to be that differences among variances are only sampling phenomena, the ordinary test of the hypothesis, $\sigma_m = 0$, being valid. If that isn't reasonable, the experiment may be divided into two or more parts, each with homogeneous variance, and

each part analyzed separately. Comparison of two means with different variances would be made as in section 4.6.

10.14—Summary. Analysis of variance has been applied to two or more groups of either the same or different sizes. The analysis furnishes estimates of population variances, (i) of individuals, (ii) of group means, and (iii) of pertinent components. It also provides for the F -tests of various null hypotheses; of $\sigma_m = 0$ if σ is the same in all groups, and more generally of a common normally distributed source of all subsamples. Chi-square is used to test the hypothesis of homogeneous variance. Finally, analysis of variance is a convenient method of computing intraclass correlation.

The mathematical model for the analysis of variance of groups is this: the groups shall be random samples from normally distributed populations. The probabilities associated with F are only slightly disturbed by moderate departures from normality.

The mean square for individuals within subsamples is a valid estimate of experimental error only if the experiment is conducted so that there would be no intraclass correlation in uniformity trials. If lots are kept in separate places their environments are likely different, with the consequent underestimate of experimental error and exaggeration of significance.

REFERENCES

1. M. S. BARTLETT. Supplement to the Journal of the Royal Statistical Society, 4:137 (1937).
2. F. L. ENGLENDOW and G. UDN YULE. The Principles and Practice of Yield Trials, Empire Cotton Growing Corporation, London (1926).
3. R. A. FISHER. International Mathematical Conference, Toronto (1924).
4. R. A. FISHER. Statistical Methods for Research Workers. Oliver and Boyd, Edinburgh (1925–1944).
5. R. A. FISHER and F. YATES. Statistical Tables, Oliver and Boyd, Edinburgh (1938–1943).
6. T. ROY HANSBERRY and CHARLES H. RICHARDSON. Iowa State College Journal of Science, 10:27 (1935).
7. J. A. HARRIS. Biometrika, 9:446 (1913).
8. M. T. JENKINS. Journal of Agricultural Research, 39:677 (1929).
9. BELLE LOWE. Data from the Iowa Agricultural Experiment Station (1935).
10. P. C. MAHALANOBIS. Indian Journal of Agricultural Science, 2:694 (1932).
11. HORATIO H. NEWMAN, FRANK N. FREEMAN and KARL J. HOLZINGER. Twins. The University of Chicago Press (1937).
12. E. S. PEARSON. Biometrika, 23:114 (1931).
13. F. B. SMITH. Data from the Iowa Agricultural Experiment Station (1931).
14. GEORGE W. SNEDECOR. Analysis of Variance and Covariance. Collegiate Press, Inc., Ames (1934).

Chapter 11

ANALYSIS OF VARIANCE WITH TWO OR MORE CRITERIA OF CLASSIFICATION

11.1—The use of knowledge about outcome. The investigator's ability to forecast differential performances among his experimental units has been ignored in the foregoing chapter. The random allotment of individuals to groups is appropriate if there is no knowledge of outcome. But in many cases it is known that units treated alike will behave differently: young male rats gain weight faster than females in identical environments, while plants in one section of a greenhouse bench produce less flowers than replicates in other sections. If units of both kinds are assigned to a lot for the purpose of evaluating the effect of a treatment the experimental error will contain not only the unpredictable variation due to sampling but also an increment from the predictable behavior of the individuals. It has been emphasized earlier in this text that experiments can be designed to eliminate from error the portion of variation which can be anticipated. For two treatments the pairing of chapter 2 was effective. We are now to learn corresponding methods for three or more treatments.

11.2—Randomized blocks on a greenhouse bench. In the soybean experiment of section 6.2 it was indicated that of the 4 plants measured in any particular week one was chosen at random from each of 4 blocks. The block consisted of 5 plants in adjacent pots: these were arranged on the bench so as to warrant the assumption that the plants would grow alike. The layout is shown in figure 11.1.

Though the 5 pots of any one block were placed so that environmental conditions (temperature, air currents, humidity, light) were as nearly identical as possible, these conditions were expected to be different for the several blocks: that they were can be seen by comparison of the heights in, for example, the fifth week. The rejection of this block variation from experimental error is the object of the statistical method now to be explained.

For convenience of examination as well as to facilitate computation, the soybean data are assembled (table 11.1) by block and date. The randomization is no longer evident, but the weekly heights and the block differences are brought into prominence. Sums of squares corresponding to these two sources of variation are to be separated from the total by means of analysis of variance.

Computation follows the pattern already familiar. First, the sum of

Block 1	$\frac{5}{44}$	$\frac{2}{18}$	$\frac{4}{38}$	$\frac{3}{26}$	$\frac{1}{4}$
Block 2	$\frac{4}{36}$	$\frac{2}{19}$	$\frac{3}{25}$	$\frac{5}{43}$	$\frac{1}{3}$
Block 3	$\frac{3}{24}$	$\frac{4}{28}$	$\frac{2}{18}$	$\frac{5}{39}$	$\frac{1}{6}$
Block 4	$\frac{1}{7}$	$\frac{2}{13}$	$\frac{3}{21}$	$\frac{4}{31}$	$\frac{5}{38}$

FIG. 11.1—Five pots in each of 4 blocks. The pots were arranged on the bench in the order shown. The upper figure for each pot is the week randomly chosen for measurement, the lower is the height of the soybean plant (centimeters).

squares for the entire experiment is the sum of the squares of the 20 heights, corrected for mean:

$$(4)^2 + (3)^2 + (6)^2 + (7)^2 + (18)^2 + \dots + (38)^2 - (480)^2/20 = 3,250$$

This is entered in table 11.2 along with the associated 19 degrees of freedom.

TABLE 11.1
HEIGHTS (CENTIMETERS) OF SOYBEAN PLANTS IN 5 WEEKS.
FOUR BLOCKS OR REPLICATIONS

Block	Week					Sum	Mean
	1	2	3	4	5		
1	4	18	26	38	44	130	26
2	3	19	25	35	43	125	25
3	6	18	24	28	39	115	23
4	7	13	21	31	38	110	22
Sum	20	68	96	132	164	480	
Mean	5	17	24	33	41		24

TABLE 11.2
ANALYSIS OF VARIANCE OF SOYBEAN HEIGHTS IN RANDOMIZED BLOCKS EXPERIMENT

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Total	19	3,250	
Weeks	4	3,120	780.
Blocks	3	50	16.67
Remainder (discrepance)	12	80	6.67

Next, there is the sum of squares for the weekly means,

$$\frac{(20)^2 + (68)^2 + (96)^2 + (132)^2 + (164)^2}{4} - \frac{(480)^2}{20} = 3,120$$

This is $4s_y^2$ as computed in table 6.3, 4 being the number of plants measured each week, and corresponds to the sum of squares for sub-samples in chapter 10.

There is now a third sum of squares for block means:

$$\frac{(130)^2 + (125)^2 + (115)^2 + (110)^2}{5} - \frac{(480)^2}{20} = 50$$

Notice the divisors in the denominators: as usual, these are the numbers of heights whose sum is squared in each term of the numerator.

In table 11.2 it is clear that the degrees of freedom for weeks and blocks fall short of the total, and that the same is true for sums of squares. The nature and utility of the remainder, *discrepance*, is to be discussed at some length.

The growth of the plants is reflected in the large mean square for weeks: most of the variation among the 20 plants is due to normal increase in height. The much smaller mean square for blocks is attributable to the differences among rates of growth in the several positions on the bench. As for the mean square called discrepancy, we shall find that it is associated with the sampling variation ordinarily thought of as experimental error, and that in this experiment it estimates the same variance as did $s_{y \cdot x}^2$ of chapter 6.

For those who prefer symbolical summaries, the data, computing instructions and analysis of variance in table 11.3 will be useful.

11.3—The nature of discrepancy. In a two-way table like 11.1 each entry is characterized by two criteria, one determining the rows (blocks) and the other the columns (weeks). From the border means an expected value may be calculated for each cell of the table, and this may then be contrasted with the observed entry in a manner analogous to that of

TABLE 11.3
SYMBOLICAL REPRESENTATION OF A 2-WAY TABLE WITH r ROWS AND c COLUMNS.
COMPUTING INSTRUCTIONS AND ANALYSIS OF VARIANCE

Row	Column				Row Total
	1	2	...	c	
1	X_{11}	X_{12}	...	X_{1c}	$S_{1.}$
2	X_{21}	X_{22}	...	X_{2c}	$S_{2.}$
...	...	...	X_{ij}	...	$S_{i.}$
r	X_{r1}	X_{r2}	...	X_{rc}	$S_{r.}$
Column total	$S_{.1}$	$S_{.2}$	$S_{.j}$	$S_{.c}$	S

Correction term = S^2/rc

Sums of Squares

Total: $T = S(X_{ij})^2 - \text{correction}$

Rows: $R = \frac{S_{1.}^2 + S_{2.}^2 + \dots + S_{r.}^2}{c} - \text{correction}$

Columns: $C = \frac{S_{.1}^2 + S_{.2}^2 + \dots + S_{.c}^2}{r} - \text{correction}$

Discrepance: $D = T - R - C$

Analysis of Variance

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Rows	$r - 1$	R	$R/(r - 1)$
Columns	$c - 1$	C	$C/(c - 1)$
Discrepance	$(r - 1)(c - 1)$	D	$D/(r - 1)(c - 1)$
Total	$rc - 1$	$T = R + C + D$	

EXAMPLE 11.1—Here are some data simulating the soybean experiment. They are chosen for ease in computation. Analyze the variance.

Block	Age			
	1	2	3	4
1	1	3	3	9
2	6	7	8	11
3	2	5	10	7

The answer is this analysis of variance:

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Age	3	60	20
Blocks	2	32	16
Discrepance	6	24	4
Total	11	116	

sections 9.5 and 9.8. The expected value is determined from a kind of regression equation:

$$\hat{X} = \bar{x}_{i.} + \bar{x}_{.j} - \bar{x},$$

in which the subscripts, i , and j , designate the row and column. To illustrate, take $i = 3$ (third block) and $j = 4$ (fourth week) of the soybean data. Here $\bar{x}_{i.} = 23$, $\bar{x}_{.j} = 33$ and $\bar{x} = 24$; hence, for this cell of the table

$$\hat{X} = 23 + 33 - 24 = 32$$

Now, $X - \hat{X} = 28 - 32 = -4$ is the deviation of the regression value from that observed, and this is attributed to sampling variation. Finally, the sum of the squares of the 20 deviations is the same as discrepancy in table 11.2. All of the regression values and deviations, calculated as above, are set out in table 11.4.

TABLE 11.4
OBSERVED AND REGRESSION VALUES TOGETHER WITH DEVIATIONS, $X - \hat{X}$,
IN THE SOYBEAN EXPERIMENT

Block	Week											
	1			2			3			4		
	X	$\hat{X}$	$X - \hat{X}$	X	$\hat{X}$	$X - \hat{X}$	X	$\hat{X}$	$X - \hat{X}$	X	$\hat{X}$	$X - \hat{X}$
1	4	7	-3	18	19	-1	26	26	0	38	35	3
2	3	6	-3	19	18	1	25	25	0	35	34	1
3	6	4	2	18	16	2	24	23	1	28	32	-4
4	7	3	4	13	15	-2	21	22	-1	31	31	0

$$S(X - \hat{X})^2 = (-3)^2 + (-3)^2 + (2)^2 + (4)^2 + (-1)^2 + \dots + (-1)^2 = 80$$

The notable features of table 11.4 are: (i) The sum of the deviations in each row and also in each column is zero. This implies that the sum of the regression values in each row (or column) is the same as the sum of the observed. (ii) As already noted, the sum of the squares of the

deviations is identical with the sum of squares for discrepancy, table 11.2.

(iii) The regression values form a pattern of differences that is the same for every column. As examples, look at columns 1 and 3:

Column 1	Column 3
$7 - 6 = 1$	$26 - 25 = 1$
$6 - 4 = 2$	$25 - 23 = 2$
$4 - 3 = 1$	$23 - 22 = 1$

The succession of differences, 1, 2, 1, is the same in all the 5 columns. A similar statement is true for rows, the pattern of differences being -12 , -7 , -9 . (iv) The discrepancy is caused by the failure of the observed values to follow such patterns of differences. This fact can be used to form a rough estimate of discrepancy in 2-way tables; observe the differences in the several columns—if they follow similar patterns, the discrepancy is small. (v) It is evident from (i) and (iv) that analysis of variance of the regression values would result in the same sums of squares for weeks and blocks as those in table 11.2, but in zero discrepancy. (vi) Because the sum of the c deviations in every row must be zero, only $c - 1$ of them are independent—the same argument that was used in

EXAMPLE 11.2—A set of data arranged for easy computation is the following:

Row	Column		
	1	2	3
1	6	5	4
2	15	10	8
3	15	15	12

Analysis of Variance

Source of Variation	Degrees of Freedom	Mean Square
Rows	2	63
Columns	2	12
Discrepance	4	2.5

Calculate the sum of squares for discrepancy by use of the deviations from regression.

EXAMPLE 11.3—In the foregoing example deduct 4 from every entry, then recalculate the analysis of variance. As you may remember, this type of coding does not affect variance (section 5.4). Do the computation again after subtracting the mean, 10, from each datum: this kind of coding is often convenient if a calculating machine is not handy.

EXAMPLE 11.4—Turn forward to table 11.5, calculate deviations from regression values in each of the 20 cells and verify the fact that the sum of their squares is 1,212.

EXAMPLE 11.5—There is no better practice in the mechanics of these 2-way tables than constructing one for easy calculation. Try it. You will find that it is well to start with deviations from mean.

EXAMPLE 11.6—Arrange a table similar to that of example 11.1 but satisfying this condition; the sum of squares for blocks shall be zero.

EXAMPLE 11.7—Make up another table having all its variation in discrepancy: the sums of squares for both rows and columns equal to zero.

EXAMPLE 11.8—If you are interested in the mathematics, you will find it profitable to verify the addition theorem for sums of squares,

$$Sx^2 = cS\bar{x}_{i.}^2 + rS\bar{x}_{.j}^2 + Sd^2,$$

where x , $\bar{x}_{i.}$ and $\bar{x}_{.j}$ are deviations from $\bar{x}$, and d is deviation from the regression value. Start with the equation,

$$\hat{X} = \bar{x}_{i.} + \bar{x}_{.j} - \bar{x} + d$$

sections 2.10 and 9.8. Similarly, there are $r - 1$ independent deviations (or degrees of freedom) in every column. Altogether, then, the degrees of freedom for the rc deviations are $(r - 1)(c - 1)$.

In following sections we shall make various statements about probability. These are all based on the hypothesis that the deviations in the 2-way table constitute a random sample from a normally distributed population with $m = 0$. The mean square for discrepancy is an estimate of the variance, σ^2 , in this population.

The hypothesis of independence of the deviations implies, among other things, that in the population there is zero correlation between these deviations in any pair of rows or columns. Of course, if you calculate some of these correlations in table 11.4, you will find plenty of sampling variation from zero; but the values will tend to center on zero in the customary distribution of small sample correlations. The hypothesis of independence of deviations has been misunderstood by some, apparently, because the statement is often made that analysis of a 2-way table is not valid if the original entries are correlated. A glance at the regression values in table 11.4 will convince you that they are perfectly correlated, so that a high correlation is to be expected among the X 's (remember common elements, section 7.5). One object of the analysis of variance is to segregate the correlated variation in the rows and columns from the independent part which estimates σ^2 .

11.4—Mean squares as estimates of population variances. Tests of null hypotheses. An easy way to visualize the facts about estimates is to outline a 2-way table then fill the cells by random drawing from a normal distribution like table 3.1. That is the way table 11.5 was made, with slight modifications to insure ease of calculation. This particular random sample was selected from a number of others because it illustrates the average unusually well.

The computations proceed as in section 11.2, the row and column sums being easily supplied. Following the argument of section 10.2 it may be

TABLE 11.5
ENTRIES DRAWN AT RANDOM FROM THE PIG GAINS OF TABLE 3.1

Outcome Level	Treatments				Means
	1	2	3	4	
1	15	31	30	20	24
2	22	11	26	45	26
3	37	30	21	36	31
4	18	31	34	49	33
5	33	37	44	30	36
Means	25	28	31	36	30

Analysis of Variance

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Treatments	3	330	110
Outcome levels	4	392	98
Discrepance	12	1,212	101
Total	19	1,934	101.8

said that each of the four mean squares is an estimate of the population variance, known to be $\sigma^2 = 100$. The first three are independent of each other, while the fourth is a weighted mean of these first three.

Table 11.5 may be looked upon as simulating a uniformity trial with swine. Each dummy treatment is applied to a lot of 5 animals, but in an actual experiment these would not be assigned to the lots wholly at random. Instead, the animal husbandman's knowledge of outcome would be utilized. Four pigs would be selected for level 1 because they are judged capable of gaining at the same rate if treated alike. These 4 would then be randomly assigned to the treatments. Thus, the quadruplet at each outcome level is an expansion of the pair in the tobacco virus experiment of section 2.13: many treatments are now possible instead of only two. Experimental error is evaluated by discrepancy—the failure of the 4 pigs in any level to gain in accord with the pattern set by the border means—just as the error in the two-treatment design was a measure of the failure of the pair differences to be identical with the mean difference.

Beginning in section 10.6, we learned that mean squares may be estimates of several components of variance. The same is true in 2-way tables, as indicated in the first part of table 11.6. One new feature appears—the row and column components are independent of each other, neither component appearing in the other mean square. Note: the coefficient of the row component, σ_R^2 , is equal to the number of columns, and *vice versa*.

TABLE 11.6
POPULATION COMPONENTS OF VARIANCE ESTIMATED BY
MEAN SQUARES IN A 2-WAY TABLE

Source of Variation	Mean Square Is an Estimate of	Sample Statistics in	
		Table 11.5	Table 11.2
Row (R)	$\sigma^2 + c\sigma_R^2$	$c=4, s_R^2 = \frac{98-101}{4} = -0.75$	$c=5, s_R^2 = 2.00$
Column (C)	$\sigma^2 + r\sigma_C^2$	$r=5, s_C^2 = \frac{110-101}{5} = 1.80$	$r=4, s_C^2 = 193.33$
Discrepance	σ^2	$s^2 = 101$	$s^2 = 6.67$

Note: c and r are the numbers of entries in the rows and columns respectively.

The computations of the components of variance, much the same as those learned in the previous chapter, are applied in the other two parts of the table to the mean squares of tables 11.5 and 11.2. As would be expected in the uniformity data of the former, there are only small components for both rows and columns, and these represent no more than accidents of sampling. As a contrast, the component for age in the soybean experiment of table 11.2, $s_C^2 = 193.33$, reflects the chief source of variation, the growth of the plants.

You have doubtless already guessed that the hypotheses, $\sigma_R^2 = 0$ and $\sigma_C^2 = 0$, are those tested by F . As an example, the test of the (wholly unrealistic) hypothesis that there is no change of soybean height with age is

$$F = \frac{s^2 + rs_C^2}{s^2} = \frac{\text{mean square for treatments}}{\text{mean square for discrepancy}} \\ = 780/6.67 = 117, d.f. = 4 \text{ and } 12$$

Significance is obvious. There may be interest, also, in learning if the block differences are attributable to sampling from a uniform population environment. $F = 16.67/6.67 = 2.50$, $d.f. = 3$ and 12 , $F_{.05} = 3.49$: the evidence against this null hypothesis is not convincing.

EXAMPLE 11.9—Hoffman (11) and Lowe (14) investigated the flavor of the breast muscle of Buff Orpington poultry subjected to the following 4 treatments:

- Drawn 2 hours after killing, then frozen and stored.
- Frozen 2 hours after killing, stored 10 days, thawed, drawn, refrozen, and stored.
- Frozen 2 hours after killing, stored 30 days, thawed, drawn, refrozen, and stored.
- Frozen 2 hours after killing, stored 90 days, thawed, drawn, and refrozen.

The chief distinction was in the period elapsing between killing and drawing. All were stored 90 days before cooking started. There were 7 replications of the 4 treatments—28 birds in all. Four judges scored the flavor.

Owing to parallel treatments not here reported, only one chicken could be cooked and scored per day. It was anticipated that the flavor might change during this 28-day period, hence one replication was completed each 4 days, the 4 treatments being taken in random order during each block of days. The sums of the 4 scores per bird are shown in the table.

Treatment	Block of 4 Days						
	1	2	3	4	5	6	7
<i>a</i>	20	18	16	17	16	14	16
<i>b</i>	18	18	16	20	15	10	16
<i>c</i>	20	18	16	18	18	15	13
<i>d</i>	20	16	20	17	17	14	19

Show that, for treatment, $F = 0.80$. There is no evidence that the flavor was affected by the time elapsing before drawing. For blocks, $F = 4.97$. Clearly the flavor did decline during the 28-day period of testing.

EXAMPLE 11.10—The following are average daily gains of 4 lots of swine receiving different rations (20). The animals were divided into 10 outcome groups. It was thought that the four in each group would gain at equal rates if fed the same.

Outcome Group	Ration Number				Sum
	1	2	3	4	
1	1.40	1.31	1.40	1.96	6.07
2	1.79	1.30	1.47	1.77	6.33
3	1.72	1.21	1.37	1.62	5.92
4	1.47	1.08	1.15	1.76	5.46
5	1.26	1.45	1.22	1.88	5.81
6	1.28	0.95	1.48	1.50	5.21
7	1.34	1.26	1.31	1.60	5.51
8	1.55	1.14	1.27	1.49	5.45
9	1.57	1.25	1.22	1.77	5.81
10	1.26	1.00	1.36	1.27	4.89
Sum	14.64	11.95	13.25	16.62	56.46
Mean	1.464	1.195	1.325	1.662	

One pig from each outcome group was randomly assigned to each ration. So far as the animal husbandman's judgment is good this scheme insures a balance of high and low gainers in the lots. On the other hand, since every pig has the same chance of receiving any one ration, there is no bias in either the means or the experimental error. Analyze the variance as follows:

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Total	39	2.2309	
Outcome groups	9	0.4074	0.0453
Rations	3	1.1986	0.3995**
Discrepance (error)	27	0.6249	0.0231

For rations, $F = 0.3995/0.0231 = 17.3$, $F_{.01} = 4.60$. In the test of outcome groups, $F = 0.0453/0.0231 = 1.96$, $F_{.05} = 2.25$. The increase in efficiency over random assignment to lots will be evaluated in example 11.15.

EXAMPLE 11.11—Try computing the analysis of variance of table 11.5 by use of deviation of both items and means from the overall mean, 30. Remember that the variance of items, estimated from means, is k times the variance of the means: for row means, $k = 4$; for columns, 5.

EXAMPLE 11.12—Calculate the sum of squares for columns in example 11.11 by taking the deviations of the column sums, 125, 140, 155, and 180 from their mean, 150. You will observe that the variance of items is one-fifth the variance of sums. It is well to keep in mind the relations,

$$k(\text{variance of means}) = \text{mean square} = (\text{variance of sums})/k$$

EXAMPLE 11.13—The facts of the two preceeding examples are hard for the beginner to keep in mind. In tables 10.11, 10.14, and 10.16 the contrast between mean square and variance was emphasized. The following statements about table 11.6 may help: (i) The variance of means, in the rows for example, is

$$(s^2 + cs_R^2)/c = s^2/c + s_R^2$$

(ii) The variance of the row sums is

$$c(s^2 + cs_R^2) = cs^2 + c^2s_R^2$$

11.5—Analysis of variance applied to 2 treatments in randomized blocks. The comparison of individuals in section 2.13 is a special case of the 2-way table, and may be thought of as a randomized blocks experiment with 2 treatments. The tobacco virus data of table 2.2 are copied in table 11.7. Variance is then analyzed by the methods of the present chapter.

TABLE 11.7
NUMBERS OF LESIONS ON HALVES OF 8 TOBACCO LEAVES.
TWO TREATMENTS REPLICATED ON 8 PLANTS
Copied from table 2.2

Plants	Treatments		Difference	Sum
	X_1	X_2	$X_1 - X_2$	$X_1 + X_2$
1	9	10	-1	19
2	17	11	6	28
3	31	18	13	49
4	18	14	4	32
5	7	6	1	13
6	8	7	1	15
7	20	17	3	37
8	10	5	5	15
Sum	120	88	32	208
Mean	15	11	4	13

Analysis of Variance

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Treatments	1	64	64
Plants	7	575	82.2
Discrepance	7	65	9.29
Total	15	704	

Test for treatments: $F = 64/9.29 = 6.89 = (2.63)^2 = t^2$ (section 2.13)

The sum of squares for treatments can be got in the usual fashion,

$$\frac{(120)^2 + (88)^2}{8} - \frac{(208)^2}{16} = 64,$$

or by the special formula mentioned in section 10.5,

$$\frac{(120 - 88)^2}{16} = 64$$

As regards discrepancy, its variance may be inferred from that of the differences, 18.57, in table 2.2. Since the variance of differences is an estimate of $2\sigma^2$, then σ^2 is estimated by $18.57/2 = 9.29$. The two methods of computation, therefore, lead to the same statistic.

As has been noted before, if $n_1 = 1$ d.f. then $t^2 = F$ as indicated in the last line of table 11.7.

The method of the present chapter leads to one result not available in chapter 2: the estimate of variance, 82.2, based on plant means. This large and highly significant mean square is the justification of the randomized blocks design—the investigator circumvented the variation from plant to plant by trying the preparations on halves of the same leaf.

The efficiency of this experiment was discussed in section 3.10, but reference had to be made to the present one for more accurate evaluation. Yates (28) gave an easily remembered rule for reconstructing the table of analysis of variance so as to arrive at an unbiased estimate of the variance as it would be if the sampling of the half leaves on the 8 plants were random: substitute the error variance, 9.29, for the mean squares of both error (7 d.f.) and treatments (1 d.f.), then calculate the corresponding sum of squares, $(8)(9.29) = 74.3$. In table 11.8, this is added to the sum of squares for plants (blocks) and the new total divided by the total degrees of freedom. Thus, the information in this experiment as compared to the random groups of chapters 4 and 10 is 466%.

TABLE 11.8

RECONSTRUCTION OF TABLE 11.7 TO GET ESTIMATE OF VARIANCE IN RANDOM SAMPLE. BLOCK SUM OF SQUARES AND ERROR MEAN SQUARE, IN BOLDFACE TYPE, COPIED FROM PREVIOUS TABLE.

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Plants (blocks)	7	575	
Error + treatments	8	74.3	9.29
Total (estimated random)	15	649.3	43.3

Efficiency of blocks as compared to random sampling = $43.3/9.29 = 466\%$

This is the last of the series of sign posts that have guided you to the easier portions of the text. The most used methods of biological statistics have been presented. Any of the following chapters may now be selected for study, or you may wish to examine more in detail some of the earlier sections omitted.

EXAMPLE 11.14—In example 2.28 you compared the gains in weight of 10 pairs of litter mates (swine). Analyze the variance of the 20 gains, as follows:

Source of Variation	Degrees of Freedom	Mean Square
Treatment	1	14.45
Litter	9	6.05
Discrepance	9	3.56

For treatments, compare F with f^2 . Compare the mean square for discrepancy with the variance of the differences, 7.12, in the earlier example.

EXAMPLE 11.15—Show that the experiment of example 11.10 had the efficiency 122% as compared to the random assignment of animals to the treatments. This means that between 12 and 13 pigs would have been required in each random group to achieve the same accuracy as was got by 10 outcome groups.

11.6—Randomized blocks in a field trial. This design has been widely used in field plot experimentation. The fact that neighboring plots tend to yield alike makes it desirable to try each treatment (or variety) on one of a block of plots lying as closely together as feasible. Uniformity of fertility throughout the block is the ideal.

The block is divided into a number of plots equal to the number of treatments. The treatments are then assigned to the plots at random to insure to each the equality of opportunity to occupy any superior position. To equalize the hazard of position in the experiment as a whole, two or more blocks or replications are laid out, each containing all the treatments. In field experiments from 3 to 10 replications are common, the number depending on the precision required.

Figure 11.2 is the plan of a field layout for testing 4 strains of Gallipoli

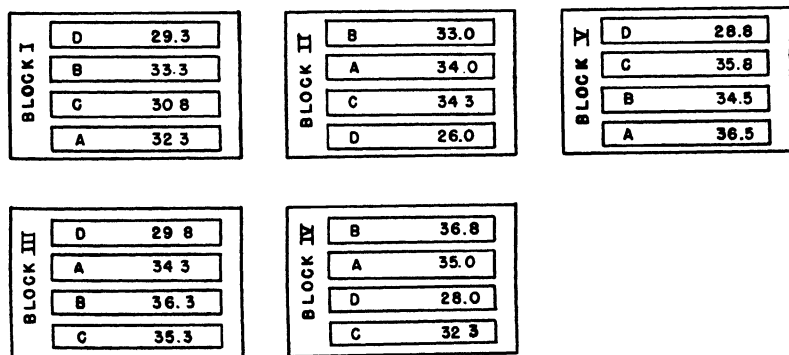


FIG. 11.2—Field plan of an experiment with four strains, A, B, C, and D, of Gallipoli wheat. Yields are in pounds per plot.

wheat in each of 5 blocks (9). Narrow plots extending the length of the block are preferable: there may be considerable variation in fertility from end to end of the block, but the aim is to have this affect all the plots alike. The blocks should not be too narrow, but should approach squareness.

The data together with the analysis of variance are recorded in table 11.9. Notice the coefficient of variation, $C = s/\bar{x} = \sqrt{2.19}/32.8$, only

TABLE 11.9
YIELDS OF FOUR STRAINS OF GALLIPOLI WHEAT PLANTED IN FIVE RANDOMIZED BLOCKS
Pounds per Plot

Strain	Block					Sum	Mean
	1	2	3	4	5		
<i>A</i>	32.3	34.0	34.3	35.0	36.5	172.1	34.4
<i>B</i>	33.3	33.0	36.3	36.8	34.5	173.9	34.8
<i>C</i>	30.8	34.3	35.3	32.3	35.8	168.5	33.7
<i>D</i>	29.3	26.0	29.8	28.0	28.8	141.9	28.4
Sum	125.7	127.3	135.7	132.1	135.6	656.4	32.8

Analysis of Variance

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Blocks	4	21.46	5.36
Strains	3	134.45	44.82**
Error	12	26.26	2.19
Total	19	182.17	

4.5%, indicating unusually accurate field plot experimentation. Another interesting feature is the similarity of the first three means. One suspects that the highly significant F , indicated by the double asterisk on the treatment mean square, is attributable largely to the small yield of *D*. It has been noted before (section 10.4) that while exact tests of differences between pairs of treatment means were not planned in the design of this experiment, adequate information springs from two sources: (i) the array of means, each an unbiased estimate of strain yield; and (ii) the fiducial limits of these means. The latter (section 3.8) are calculated from

$$t_{.05; 2} = 2.179\sqrt{2.19/5} = \pm 1.4 \text{ pounds per plot}$$

The upper limit for *D* is $28.4 + 1.4 = 29.8$ pounds, while the lower for the next mean in the array is $33.7 - 1.4 = 32.3$ pounds. Since there is no overlapping of the intervals, the evidence is quite convincing that the means for *C* and *D* are not samples from the same population. On the

contrary, the intervals for the other 3 means indicate no distinctions among the corresponding populations.

An experiment like this might have been designed to test three new strains, *A*, *B*, and *C*, against the standard, *D*. In that case, the 3 degrees of freedom for strains would naturally fall into two parts; 2 *d.f.* for differences among the 3 new strains and 1 for the contrast, old vs. new. The sum of squares for the first part is

$$\frac{(172.1)^2 + (173.9)^2 + (168.5)^2}{5} - \frac{(514.5)^2}{15} = 3.02$$

For the single degree of freedom, the sum of squares is (section 10.8)

$$\frac{(514.5)^2}{15} + \frac{(141.9)^2}{5} - \frac{(656.4)^2}{20} = 131.43$$

Note that the sum of the two parts is the strain sum, 134.45. For this design the analysis in table 11.10 would be appropriate.

TABLE 11.10

ANALYSIS OF HYPOTHETICAL EXPERIMENT DESIGNED TO CONTRAST 3 NEW STRAINS, *A*, *B*, AND *C*, WITH THE STANDARD, *D*. THE DATA ARE THOSE IN TABLE 11.9.

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Strains <i>A</i> , <i>B</i> , and <i>C</i>	2	3.02	1.51
Old vs. new	1	131.43	131.43**
Error (table 11.9)	12		2.19

The moral is plain: exact tests of significance are available for comparisons built into the experimental structure, but no definite probability statements can be made about contrasts suggested by the data. Further information about individual comparisons will be found in chapter 15.

Some people still question the advisability of random arrangements for field plots. There are many arguments pro and con, and they are all complicated. One obstacle to clear understanding is the unlimited number of nonrandom (that is, systematic) arrangements: each must be considered separately. As examples, the "knight's move" was investigated by Tedin (23), and the "half drill strip" by Gosset ("Student") (22). For a more general discussion, see (8).

The simplest and least defensible of systematic layouts is the same succession, *ABCD* for example, in all the blocks. If there is any uniform gradient in fertility the result may be that, in every application, *A* is on better soil than the other treatments. This is not common, but I have seen it happen. Another difficulty with this scheme is that experimental error is biased downwards with a consequent exaggeration of significance. To see why, recall two facts: (i) plots equally distant tend to differ by the

same amount; and (ii) the estimate of error is based on the failure of yield differences, such as $A - C$, to be identical in the several blocks. Hence, if A and C are the same distance apart in every replication, the differences are likely to be more alike than if the 2 treatments are randomized. Another generally applicable argument is that A and B , being contiguous, are subject to less error than are A and D : discrepancy, an average of these errors, is not appropriate for either comparison. Randomization tends to produce an unbiased error equally suitable for testing all differences, and it is a necessary condition for exactness in the probabilities associated with fiducial statements and tests of null hypotheses.

The theory of experimentation on which this text is founded specifies erroneous inferences in a predetermined percentage (5%, for instance) of trials. Any particular random arrangement of plots applied to an area chosen for the experiment may turn out to be the worst possible. The experimental approach is not designed for him who makes a single assay: it is only the average of many trials that is guaranteed to approach the true value. During a life of research the investigator is assured that, provided he makes a fresh randomization for every trial, he will be led to correct probability statements in the chosen fraction of his experiments.

The reader who wishes to learn more about randomized blocks may consult references (7) section 48, (8) chapter 4, and (5) chapter 2. There, more complicated designs also are discussed.

Accidents often result in the loss of data in some of the cells of a 2-way table. Crops may be destroyed, animals may die, or errors may be made in recording. One missing item can be supplied with the aid of a formula proposed by Allan and Wishart (1) and modified by Yates (26):

$$X = \frac{tT + bB - S}{(t-1)(b-1)}$$

where t = number of treatments
 b = number of blocks
 T = sum of items with same treatment as missing item
 B = sum of items in same block as missing item
 S = sum of all observed items

As an example, suppose the item 29.3 pounds were missing from table 11.9, block 1, strain D . Omitting this item from the sums for block 1, strain D and total,

$$B = 96.4, T = 112.6, S = 627.1, b = 5, t = 4$$

Then, $X = [(4)(112.6) + (5)(96.4) - 627.1]/(3)(4) = 25.4$

This value would be entered in the table as the yield of the missing plot. The analysis of variance would then proceed as usual, with this one modification: the number of degrees of freedom for *total* and *error* would each be decreased by unity, becoming 18 and 11 instead of 19 and 12.

The theory is that X is so determined that the *error* sum of squares is as small as possible. This means that the supplied item is its own expected value (section 11.3), its deviation from expected being zero. Nevertheless, the deduction of a degree of freedom for the missing datum makes the estimate of error unbiased. But the treatment mean square is biased upward, so that the F -value tends to be exaggerated somewhat. Hence, it would be wise to think of a significant F as nearer to the 1% than to the 5% point.

If two or more data are missing, very good results can be got by an iterative method. If X_1 and X_2 denote two missing values, first enter in the table a reasonable value for X_2 (the experimental mean, for example, or the mean of the treatment and block containing X_2), then substitute in the formula for X_1 values of T , B , and S which include X_2 . Now set in the table the calculated X_1 , remove X_2 , recompute T , B , S , and substitute in the formula to obtain a better approximation to X_2 . Repeat the process for an improved X_1 . After a few cycles the new estimate will be practically the same as the old. The data are then ready for analysis of variance with two degrees of freedom deducted from *total* and *error*.

For practice, turn to example 11.2 and assume that data are missing from row 2, column 2 (X_1) and from row 3, column 1 (X_2). Enter 18 for X_2 because both row and column totals are large. The formula gives $X_1 = 11.5$. Supply this value in row 2, column 2, eliminate the entry for X_2 , then again substituting in the formula calculate $X_2 = 16.9$. Repetition of the process results in $X_1 = 11.8$ and $X_2 = 16.8$. The analysis of the resulting variance is,

Source of Variation	Degrees of Freedom	Mean Square
Rows	2	72
Columns	2	16
Discrepance	2	3.2

The ease of analyzing experiments with missing data may deceive you into thinking that such losses are unimportant. Actually, the loss of 1 plot in 12 means the loss of one-twelfth of the information planned in the experiment. The only objective of the foregoing methods is to salvage as much as possible of the remaining eleven-twelfths.

EXAMPLE 11.16—In the Gallipoli wheat trial, calculate the efficiency of the randomized blocks as compared to unstratified random sampling. Ans. 131%. Stratification produced greater precision than another replication would have accomplished in a completely random arrangement.

EXAMPLE 11.17—In a cooperative seed treatment trial (16) the numbers of plants emerging from 100 planted soybean seeds, Kanro variety, were reported for

each plot of a randomized blocks experiment conducted by the South Carolina Truck Experiment Station:

Block	Treatments				
	Untreated	Arasan	Spergon	Semesan, Jr.	Fermate
1	92	98	96	97	91
2	90	94	90	95	93
3	88	93	91	91	95
4	87	89	92	90	95
5	89	95	90	94	97

Analyze the variance.

Source of Variation	Degrees of Freedom	Mean Square
Treatment	4	20.95*
Blocks	4	12.45
Error	16	5.412

If you compare the mean of treated with the mean of untreated plots you will get $93.3 - 89.2 = 4.1$ plants, with the corresponding mean square, 67.24. Hence, $F = 12.42^{**}$.

EXAMPLE 11.18—Following are yields of corn (bushels per acre at 15% moisture) on Marshall silt loam in Iowa (1944):

Block	Plowed	Listed	Sub-surface Tilled	Disked	Total
1	102	93	75	78	348
2	94	96	59	71	320
3	...	88	73	66	227
Total	196	277	207	215	895

Supply a number for the missing yield (Ans. 95 bushels per acre), then analyze the variance. Ans. F is 17.5 with $d.f. = 3$ and 5.

EXAMPLE 11.19—In the foregoing example the yield, 102 bushels per acre, was estimated from the undamaged portion of a plot whose stand had been partly destroyed during cultivation. Eliminate this datum from the table above, then supply yields for two missing plots by the recursion method. Ans. In block 1, 100.7 bushels per acre, and in block 3, 94.4. The analysis of variance results in the following degrees of freedom and mean squares:

Treatment	3	588.49
Error	4	43.17

EXAMPLE 11.20—After supplying the two missing data of example 11.19, calculate regression values for the 12 yields (section 11.3). The data supplied should be approximately equal to their own $\bar{X}$'s, their deviations being zero. These missing plots contribute nothing to error, hence the loss of a degree of freedom for each.

EXAMPLE 11.21—The mathematically inclined will wish to derive the formula for a missing item. Substitute X for the yield in the blank cell, then carry through the

computations in the usual way. At the end, the sum of squares for error will contain terms in both X and X^2 . Calculate the value of X which minimizes the error.

EXAMPLE 11.22—Any systematic arrangement is one of the possible random arrangements. Suppose your drawing led to $ABCD$ in every block, what would you do?

11.7—The latin square. This experimental design perfected by Fisher (7) is especially well adapted to the field trial of usually not more than eight treatments or varieties. Interest in the design is by no means so limited, however, since it has remarkable theoretical as well as practical implications (8, chapters 5 and 7). The number of replications must be the same as the number of treatments. The field, which should be a rectangle somewhat approximating squareness, is divided into two sets of blocks at right angles. One set is usually designated *rows*, the other *columns*. If there are k rows, then there are k columns and k^2 plots. Upon the random assignment of the treatments to the plots there is placed this restriction: each treatment must occur once and only once in each row and each column.

The design is illustrated by table 11.11 (13) where the treatments

TABLE 11.11
YIELDS (GRAMS) OF PLOTS OF MILLET ARRANGED IN A LATIN SQUARE
Spacings: A, 2-inch; B, 4; C, 6; D, 8; E, 10

Row	Column					Sum
	1	2	3	4	5	
1	B: 257	E: 230	A: 279	C: 287	D: 202	1,255
2	D: 245	A: 283	E: 245	B: 280	C: 260	1,313
3	E: 182	B: 252	C: 280	D: 246	A: 250	1,210
4	A: 203	C: 204	D: 227	E: 193	B: 259	1,086
5	C: 231	D: 271	B: 266	A: 334	E: 338	1,440
Sum	1,118	1,240	1,297	1,340	1,309	6,304
Summary by Spacing						
	A: 2"	B: 4"	C: 6"	D: 8"	E: 10"	
Sum	1,349	1,314	1,262	1,191	1,188	6,304
Mean	269.8	262.8	252.4	238.2	237.6	252.2

Analysis of Variance

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Total	24	36,571	
Spacings	4	4,156	1,039
Rows	4	13,601	3,400
Columns	4	6,146	1,536
Error	12	12,668	1,056

A, B, C, D, and E are located as they were in the field. Since each row and column contains every treatment, differences among either set are attributed to soil variation. It is clear that the rows and columns should be not far from the same width in order to share alike in the soil heterogeneity. This results in desirably compact, almost square plots. Since the plots are usually small, the soil variation in them is small. Unlike the randomized block design whose blocks need not be contiguous, the latin square loses its advantages if the plots are not continuous in the field. On this account, the square is somewhat less flexible than the blocks. Also, complete loss of data for a treatment (or row or column) increases the difficulty of analysis (29). Under suitable conditions, however, the square is, on the average, about 20% more efficient than the blocks.

The data in table 11.11 result from an experiment on the effect of spacing millet plants in the row. The letters denote 2-, 4-, 6-, 8-, and 10-inch spacings. This table shows not only the planting arrangement, but also the plot yields and the sums for rows and columns. In the last two lines are the sum and mean for each spacing.

In the calculations below, leading to the analysis of variance, there is only one addition to those heretofore described, that yielding the sum of squares for the columns.

$$1. \text{ Correction: } (6,304)^2/25 = 1,589,617$$

$$2. \text{ Total: } (257)^2 + \dots + (338)^2 - 1,589,617 = 36,571$$

$$3. \text{ Spacings: } \frac{(1,349)^2 + \dots + (1,188)^2}{5} - 1,589,617 = 4,156$$

$$4. \text{ Rows: } \frac{(1,255)^2 + \dots + (1,440)^2}{5} - 1,589,617 = 13,601$$

$$5. \text{ Columns: } \frac{(1,118)^2 + \dots + (1,309)^2}{5} - 1,589,617 = 6,146$$

$$6. \text{ Remainder: } 12,668$$

The mean squares are estimates of these population variances:

Spacings	$\sigma^2 + 5\sigma_s^2$
Rows	$\sigma^2 + 5\sigma_R^2$
Columns	$\sigma^2 + 5\sigma_C^2$
Error	σ^2

The testing of null hypotheses about spacings, rows, and columns clearly offers no difficulties.

From the mean squares one concludes that the spacing effect is negligible, the variation being just that to be expected in sampling from a

single normal population of yields. But the design of the experiment included a feature not yet brought into the analysis—the regression of yield on spacing. The summary by spacing in table 11.11 shows mean plot yield decreasing with wider spacing. It appears, then, that the sum of squares for spacings, 4,156, contains a considerable portion attributable to linear regression, and that it will be informative to subdivide this sum accordingly. We shall need Sx^2 , Sxy , and Sy^2 in the manner of chapter 6. The latter is already calculated, 4,156. The other two are conveniently computed from means, then expanded to the item basis (characteristic of sums of squares and mean squares) through multiplication by $k = 5$.

$$Sx^2 = (2)^2 + \dots + (10)^2 - (30)^2/5 = 40$$

$$Sxy = (2)(269.8) + \dots + (10)(237.6) - \frac{(30)(1,260.8)}{5} = -178$$

The desired sums of squares and products, therefore, are 200 and -890 . Finally, Sy^2 is separated into the two parts: (i) attributable to regression, $(Sxy)^2/Sx^2 = (-890)^2/200 = 3,960$; and (ii) deviations from regression, $4,156 - 3,960 = 196$. The revised analysis of variance (omitting rows and columns) is in table 11.12.

TABLE 11.12
ANALYSIS OF REGRESSION OF SPACING MEANS ON WIDTH OF SPACING.
MILLET EXPERIMENT

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Regression	1	3,960	3,960
Deviations	3	196	65.7
Error (table 11.10)	12		1,056

For regression, $F = 3,906/1,056 = 3.70$, $F_{.05} = 4.75$

It now appears that most of the variation in yield is associated with regression, deviations from regression being trivial. To put it another way, the major portion of the sum of squares for spacings belongs to the single degree of freedom for regression. Even so, this regression is not significant so that there may be no relation between spacing and yield in the population. It is interesting, however, that, in the sample, the notable increase of yield per plant due to freedom from competition in the wider spacings failed to compensate for the fewer plants. Notice the sizable coefficient of variation, $C = \sqrt{1,056/252.2} = 13\%$, which may explain the non-significance of the rather striking regression.

Missing data in a latin square can be supplied in the manner explained in the preceding section. The formula is now

$$X = \frac{k(R + C + T) - 2S}{(k - 1)(k - 2)},$$

where k = number of rows (or columns or treatments)
 R = sum of items in same row as missing item
 C = sum of items in same column as missing item
 T = sum of items with same treatment as missing item
 S = sum of all observed items

For an illustration, assume that in example 11.23 the plot yield with treatment A in row 1 is missing. We have $R = 52$, $C = 43$, $T = 28$, $S = 163$, $k = 3$,

$$X = \frac{3(52 + 43 + 28) - 326}{(3 - 1)(3 - 2)} = 21.5$$

EXAMPLE 11.23—Here is a latin square for easy computation. Treatments are indicated by A , B , and C .

Rows	Columns		
	1	2	3
1	$B: 23$	$A: 17$	$C: 29$
2	$A: 16$	$C: 25$	$B: 16$
3	$C: 24$	$B: 18$	$A: 12$

The mean squares are: rows, 21; columns, 3; varieties, 93; remainder, 3.

EXAMPLE 11.24—Code the items in the latin square of the last example by subtracting some easy number like 10 or 12, then duplicate the analysis of variance.

EXAMPLE 11.25—Do the analysis of variance in example 11.23 by use of deviations from mean.

EXAMPLE 11.26—In the manner of section 11.3 insert regression values in the table of example 11.23. Each will be the sum of four parts: (i) the row mean, (ii) the column mean, (iii) the treatment mean, and (iv) minus *twice* the general mean. The sum of the 9 deviations is zero, while the sum of their squares is 6, the *remainder* in the table of analysis of variance.

EXAMPLE 11.27—You can learn a lot about latin squares by constructing one for easy computation like that of example 11.23, especially if you start with sets of deviations for row, column and variety and put in the expected numbers first.

EXAMPLE 11.28—A 7×7 latin square was laid out to test the effect of various fertilizers on the yield of potatoes. Here is the field plan with the plot yields (bushels per acre). The letters specify the treatments.

Row	Column						
	1	2	3	4	5	6	7
1	C: 344	B: 428	A: 423	D: 432	E: 449	F: 449	G: 410
2	D: 367	A: 430	F: 443	C: 387	B: 447	G: 426	E: 372
3	B: 317	C: 389	D: 425	A: 389	G: 408	E: 397	F: 384
4	E: 303	D: 393	G: 442	F: 427	C: 386	B: 358	A: 342
5	A: 291	E: 423	B: 412	G: 392	F: 364	C: 341	D: 333
6	F: 261	G: 374	E: 398	B: 347	D: 323	A: 284	C: 273
7	G: 258	F: 376	C: 404	E: 317	A: 234	D: 250	B: 233

Compute the analysis of variance:

Source of Variation	Degrees of Freedom	Mean Square
Rows	6	14,143
Columns	6	10,831
Treatments	6	1,936*
Error	30	635

11.8—More than one item in the subclasses. Some of you are asking: “Is discrepancy always a valid estimate of error? Is there some logical foundation for this in experimentation?” To the first question, the answer is, “No.” The second requires some elaboration. In the experiments hitherto presented, there is both logical and experimental justification for the use of discrepancy as error. On the logical side, randomization together with rather uniform conditions leave in discrepancy no variation except such as cannot be controlled. A long series of studies of uniformity trials, in field crops especially, have furnished the experimental evidence. In such standardized experiments discrepancy can be relied upon to measure experimental error.

If you are beginning some investigations where nothing is known of the validity of discrepancy as error, you should provide at least two and preferably three determinations of the variate in each subclassification. The method is illustrated in table 11.13, where three such determinations were made (17).

This was a greenhouse experiment to learn the effects of two kinds of soil treatments. Each combination of treatments, as for example “soil plus straw” with “carbon dioxide gas,” was laid down in three pots. To eliminate the possibility that position might affect the pots differentially the 48 pots were placed at random in the space provided in the greenhouse. Don’t overlook this precaution. If you place together three pots containing one combination of treatments, and in a second place those with another combination, the effects of position and treatment will be confounded and the three pots may be worth no more than a single determination.

TABLE 11.13

YIELD OF WHEAT IN 48 POTS. GREENHOUSE EXPERIMENT WITH TWO SERIES OF SOIL TREATMENTS, THREE POTS FOR EACH COMBINATION
Grams

Humus and Fertilizer Treatment	Pot	Chemical Treatment				Sum, 12 Pots	Yield per Pot
		None	N + O	CO ₂ Gas	Carbonic Acid		
None	1	21.4	20.9	19.6	17.6	230.2	19.2
	2	21.2	20.3	18.8	16.6		
	3	20.1	19.8	16.4	17.5		
	Sum	62.7	61.0	54.8	51.7		
Straw	1	12.0	13.6	13.0	13.3	156.7	13.1
	2	14.2	13.3	13.7	14.0		
	3	12.1	11.6	12.0	13.9		
	Sum	38.3	38.5	38.7	41.2		
Straw + PO ₄	1	13.5	14.0	12.9	12.4	160.2	13.4
	2	11.9	15.6	12.9	13.7		
	3	13.4	13.8	13.1	13.0		
	Sum	38.8	43.4	38.9	39.1		
Straw + PO ₄ + lime	1	12.8	14.1	14.2	12.0	164.6	13.7
	2	13.8	13.2	13.6	14.6		
	3	13.7	15.3	13.3	14.0		
	Sum	40.3	42.6	41.1	40.6		
Sum, 12 pots		180.1	185.5	173.5	172.6	711.7	
Yield per pot		15.0	15.5	14.5	14.4		

From the viewpoint of computation there is only one new feature. One must pool the sums of squares in the 16 subclasses. Instead of doing this directly it is easier to proceed thus:

$$1. \text{ Correction: } \frac{(711.7)^2}{48} = 10,552.44$$

$$2. \text{ Total: } (21.4)^2 + (21.2)^2 + (20.1)^2 + (12.0)^2 + \dots + (14.0)^2 - 10,552.44 = 367.15$$

3. Subclasses:

$$\frac{(62.7)^2 + (38.3)^2 + \dots + (40.6)^2}{3} - 10,552.44 = 340.87$$

$$4. \text{ Within subclasses: } 367.15 - 340.87 = 26.28$$

You can easily verify the last result by adding the 16 sums of squares you get from applying the usual computational method to each of the

subclasses. In the first subclass, for example, the sum of squares is $(21.4)^2 + (21.2)^2 + (20.1)^2 - (62.7)^2/3 = 0.98$.

The sum of squares for subclasses is now divided into three parts in the manner already familiar:

5. Humus treatments:

$$\frac{(230.2)^2 + \dots + (164.6)^2}{12} - 10,552.44 = 306.24$$

6. Chemical treatments:

$$\frac{(180.1)^2 + \dots + (172.6)^2}{12} - 10,552.44 = 9.17$$

7. Subclass discrepancy: $340.87 - (306.24 + 9.17) = 25.46$

The results are recorded in table 11.14. As will be explained below,

TABLE 11.14
ANALYSIS OF VARIANCE OF WHEAT YIELDS IN SOIL TREATMENT EXPERIMENT

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Humus treatments	3	306.24	102.08**
Chemical treatments	3	9.17	3.06*
Interaction	9	25.46	2.83**
Pots treated alike	32	26.28	0.82
Total	47	367.15	

the subclass discrepancy is known as *treatment interaction*.

The degrees of freedom for pots are worthy of note. Each subclass contributes two, making $(16)(2) = 32$ in all. The mean square, 0.82, measures the average variation of yields from pots treated alike and placed at random in the experimental space. It evaluates those uncontrolled sources of variation—seed, soil, and position—properly assigned to experimental error.

The striking feature of this experiment is the discrepancy among the subclasses. The chemicals applied to one humus treatment produced yields out of accord with those from other humus treatments. An outstanding illustration lies in the differences between successive chemical treatments in the rows "None" and "Straw + PO_4 ":

None:	1.7,	6.2,	3.1
Straw + PO_4 :	-4.6,	4.5,	-0.2

It is the failure of such pairs of differences to be alike, you remember, that produces discrepancy. Its value in this experiment indicates that

the effect of a chemical treatment is not always the same, but depends upon the humus treatment to which it is applied. The term used to describe this phenomenon is *interaction*: it is the discrepancy between experimental treatments. Its measurement and test of significance were planned in the design of this experiment. In field trials it is generally assumed with some evidence that with fairly uniform soil and in trials of varieties or treatments not obviously different there is no real interaction. The discrepancy, therefore, measures only experimental error. In the soil treatment experiment, on the contrary, there was no reason to believe that the treatments of one set would react alike with those of the other set; hence, the provision of 3 pots per combination to furnish an estimate of random variation on the greenhouse bench.

The components of variance in this kind of experimental design are:

Treatment 1 (chemical)	$\sigma^2 + 3\sigma_{12}^2 + 12\sigma_1^2$
Treatment 2 (humus)	$\sigma^2 + 3\sigma_{12}^2 + 12\sigma_2^2$
Interaction	$\sigma^2 + 3\sigma_{12}^2$
Individuals	σ^2

Evidently, interaction is tested thus:

$$F = \frac{s^2 + 3s_{12}^2}{s^2} = \frac{2.83}{0.82} = 3.45$$

The manner of testing treatments depends upon conditions under which the null hypothesis is set up. In this experiment, the hypothesis, $\sigma_1 = 0$ is to be tested on the assumption that the chemical treatments are always to be tried with this same set of humus treatments—this latter set constitutes a population, and there will be no changes in it to affect the variation of the chemical treatments. Under this condition, s^2 estimates the real error in evaluating treatment 1, and the appropriate test of $\sigma_1 = 0$ is

$$F = \frac{s^2 + 3s_{12}^2 + 12s_1^2}{s^2} = \frac{3.06}{0.82} = 3.73$$

Similarly, if the hypothesis, $\sigma_2 = 0$, is tested on the assumption of unchanging chemical treatments, then $F = 102.08/0.82$.

In some experiments to be described later (section 11.13 *et seq.*), one set of treatments, treatments 2 for example, might be considered as a random sample of possible such treatments, so that differences among treatments 1 would be affected by the particular sample of treatments 2 which happened to be drawn. Under this condition you will find that the estimate of error must be correspondingly increased, $\sigma_1 = 0$ being tested by

$$F = \frac{s^2 + k_1s_{12}^2 + k_2s_1^2}{s^2 + k_1s_{12}^2}$$

A glance at the sums for the humus treatments shows a striking contrast between those for the treated pots and that for the untreated. This

calls for separation of the sum of squares, 306.24, into two parts; one corresponding to variation among the treatment means and the second to the difference between all the straw treatments and none. The sum of squares for interaction also may be heterogeneous. For subdivision of both, it is convenient to condense the sums of the subclasses thus:

Humus	Chemical Treatment				Sum
	1	2	3	4	
Straw	117.4	124.5	118.7	120.9	481.5
None	62.7	61.0	54.8	51.7	230.2
Sum	180.1	185.5	173.5	172.6	711.7
Difference	54.7	63.5	63.9	69.2	

In steps 3 and 5 of the computation, note the modification due to unequal numbers in the subclasses:

$$3. \text{ Subclasses: } \frac{(117.4)^2 + \dots + (120.9)^2}{9} + \frac{(62.7)^2 + \dots + (51.7)^2}{3} - 10,552.44 = 333.58$$

$$5. \text{ Humus treatment: } \frac{(481.5)^2}{36} + \frac{(230.2)^2}{12} - 10,552.44 = 303.63$$

$$6. \text{ Chemical treatments: as before, 9.17}$$

$$7. \text{ Interaction: } 333.58 - (303.63 + 9.17) = 20.78$$

This interaction measures dissimilarities among the differences in the last line of the condensed table. The upward trend seems to be due to a downward trend in the no-humus sums rather than to changes in the treatment sums.

Table 11.15 contains the completed analysis of variance. It is now

TABLE 11.15
COMPLETED ANALYSIS OF VARIANCE OF THE WHEAT YIELDS IN TABLE 11.13

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Pots in subclasses	32	26.28	0.82
Chemical treatments	3	9.17	3.06*
Humus treatments:			
Three treatments	2	2.61	1.30
Treatment vs. check	1	303.63	303.63**
Interactions:			
Three humus treatments with 4 chemical	6	4.68	0.78
Humus treatment vs. humus check, with 4 chemical	3	20.78	6.93**
Total	47	367.15	

evident that the differences that may be attributed to the population are (i) those among the chemical treatment means; (ii) the difference between the means of straw treatments and no humus treatment,

$$230.2/12 - 481.5/36 = 5.8 \text{ grams;}$$

and (iii) the interaction between chemical treatments and the 2 humus treatments, straw and none.

EXAMPLE 11.29—A cooperative experiment on the effect of fertilizer resulted in the data below. The yields are in decagrams per sample of broadcast oats, the samples being $36'' \times 36''$ harvested squares, four of them from each $21' \times 52'$ plot. Soil, Carrington loam.

Block	Check	N	P	NP
1	19	18	16	13
	16	16	22	27
	18	17	18	28
	18	14	18	28
2	15	22	19	15
	16	24	20	14
	16	22	20	18
	17	20	24	14
3	21	20	24	25
	19	21	15	20
	18	20	19	24
	22	20	20	20
4	18	20	20	20
	20	22	20	18
	20	18	17	12
	18	19	20	18

Fertilizers containing nitrogen and phosphorous were applied both separately and in combination—a *factorial design*. Calculate the following analysis of variance.

Source of Variation	Degrees of Freedom	Mean Square	Variance of Plot Mean
Treatment	3	7.6	
Block	3	12.7	
Experimental error	9	31.78	7.94
Sampling error	48	7.46	

Note the large coefficient of variation, $C = 14.7\%$. This is due chiefly to field variation because the sampling error is only a small part of the experimental error. Estimate the variance of a plot mean if the sampling had been increased to 8 quadrats per plot. *Ans.* Doubling the sampling rate would decrease the experimental error by only 11.7% .

EXAMPLE 11.30—The pots of table 11.13 might have been placed on the bench in 3 blocks with the 16 treatment combinations randomized in each. In that case the 32 degrees of freedom for pots would have been subdivided in this manner:

Blocks	2
Discrepance	30

Do you think this would have been a better design?

EXAMPLE 11.31—Calculate directly the subclass discrepancy, 4.67, in the 3×4 table of sums comprising the straw and chemical treatments but omitting the humus check. Follow computing directions 3, 5, 6, and 7.

EXAMPLE 11.32—The gains in weight (grams) of six lots of male rats are shown in this table:

High Protein			Low Protein		
Beef	Cereal	Pork	Beef	Cereal	Pork
73	98	94	90	107	49
102	74	79	76	95	82
118	56	96	90	97	73
104	111	98	64	80	86
81	95	102	86	98	81
107	88	102	51	74	97
100	82	108	72	74	106
87	77	91	90	67	70
117	86	120	95	89	61
111	92	105	78	58	82

Calculate the analysis of variance as follows:

Source of Variation	Degrees of Freedom	Mean Square
Level of protein	1	3,168.3
Source	2	133.2
Interaction	2	589.0
Error	54	214.6

11.9—Proportional subclass numbers. For one reason or another the numbers of observations in the subclasses of a table like 11.13 may not be equal. When through accident, or under the limitations of an experimental technique, erratic losses of data occur, the addition theorem for sums of squares is annulled and the analysis of variance must be modified. Appropriate methods are described in following sections. Occasionally, however, the subclass numbers are proportional, as in some of the foregoing tables, causing no injury to the analysis of variance.

In table 11.16 the numbers of dressing percentages (each decreased by 70%) for the breeds and sexes represent approximately the proportions in which the animals were brought in for slaughter at the College Meats Laboratory (3). For illustration we have chosen only a small sample from the original data. All the pigs represented have weights between 200 and

TABLE 11.16
DRESSING PERCENTAGES (LESS 70%) OF 93 SWINE CLASSIFIED BY BREED AND SEX.
LIVE WEIGHTS 200-219 POUNDS

Number	Breed									
	1		2		3		4		5	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
1	13.3	18.2	10.9	14.3	13.6	12.9	11.6	13.8	10.3	12.8
2	12.6	11.3	3.3	15.3	13.1	14.4	13.2	14.4	10.3	8.4
3	11.5	14.2	10.5	11.8	4.1		12.6	4.9	10.1	10.6
4	15.4	15.9	11.6	11.0	10.8		15.2		6.9	13.9
5	12.7	12.9	15.4	10.9			14.7		13.2	10.0
6	15.7	15.1	14.4	10.5			12.4		11.0	
7	13.2		11.6	12.9					12.2	
8	15.0		14.4	12.5					13.3	
9	14.3		7.5	13.0					12.9	
10	16.5		10.8	7.6					9.9	
11	15.0		10.5	12.9						
12	13.7		14.5	12.4						
13			10.9	12.8						
14			13.0	10.9						
15			15.9	13.9						
16			12.8							
17			14.0							
18			11.1							
19			12.1							
20			14.7							
21			12.7							
22			13.1							
23			10.4							
24			11.9							
25			10.7							
26			14.4							
27			11.3							
28			13.0							
29			12.7							
30			12.6							
<i>SX</i>	168.9	87.6	362.7	182.7	41.6	27.3	79.7	33.1	110.1	55.7

Total: $N = 93$, $SX = 1,149.4$, $SX^2 = 14,785.62$

Breed sums: 1, 256.5; 2, 545.4; 3, 68.9; 4, 112.8; 5, 165.8.

Sex sums: Male, 763.0; Female, 386.4.

219 pounds: in this way any difficulties that might inhere in percentages are avoided. Let us emphasize the fact that the subclass numbers are representative of population proportions. The picture would be distorted by equal class sizes. The only modification we have introduced into the original data has been to make the numbers proportional; that is,

$$\begin{array}{l} \text{both} \quad 12:6 = 30:15 = 4:2, \text{ etc.} \\ \text{and} \quad 12:30:4:6:10 = 6:15:2:3:5 \end{array}$$

While proportional subclass numbers do not disturb the analysis of variance the computations must be slightly modified because of the unequal sized groups (section 10.8).

Here is the method:

1. Correction: $C = (SX)^2/n = (1,149.4)^2/93 = 14,205.60$
2. Total: $SX^2 - C = 14,785.62 - 14,205.60 = 580.02$
3. Subclasses: $\frac{(168.9)^2}{12} + \frac{(87.6)^2}{6} + \dots + \frac{(55.7)^2}{5} - C = 122.83$
4. Within subclasses: $580.02 - 122.83 = 457.19$
5. Breed: $\frac{(256.5)^2}{18} + \dots + \frac{(165.8)^2}{15} - C = 97.38$
6. Sex: $\frac{(763.0)^2}{62} + \frac{(386.4)^2}{31} - C = 0.52$
7. Interaction: $122.83 - (97.38 + 0.52) = 24.93$

The analysis of variance is given in table 11.17.

TABLE 11.17
ANALYSIS OF VARIANCE OF DRESSING PERCENTAGES

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Breed	4	97.38	24.34**
Sex	1	0.52	0.52
Breed-sex interaction	4	24.93	6.23
Within subclasses	83	457.19	5.51
Total	92	580.02	

EXAMPLE 11.33—Between 1934 and 1942, the Test of Performance records the Iowa Agricultural Experiment Station included 9 litters each with 1 barrow and 2 sows. The weights at 180 days of age were as follows:

Litter	Barrow	Sows	Litter	Barrow	Sows
1	225	209 217	6	222	211 193
2	180	216 194	7	175	140 133
3	146	162 118	8	207	198 186
4	252	200 234	9	189	200 179
5	134	161 161			

Calculate the analysis of variance:

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Litters	8	23,917.9	2,989.7
Sex	1	405.6	405.6
Litter $\times$ sex	8	2,838.7	354.8
Sows of same litter	9	2,299.0	255.4

Since the sample chosen for illustration represents only a small fraction of the original data, conclusions should be only tentative.

11.10—Disproportionate subclass numbers. The 2×2 table. Neither equality nor proportionality in subclass numbers is always attainable. For example, one can't control the proportion of the sexes of chicks hatched from differently treated batches of eggs. The consequences may be serious if the character measured, such as growth rate, is different in the sexes. A lot of chicks with a predominance of males will likely gain faster than one with extra females even though the treatment of the second lot is more favorable to growth. Much the same effect may be observed in table 11.18. The weighted mean of the 2 lots injected with hormone B is the smaller, whereas the unweighted mean (ignoring subclass numbers) is the greater. As we shall learn, neither pair gives an unbiased estimate of the effect of *B*.

TABLE 11.18
COMB WEIGHTS (MILLIGRAMS) OF LOTS OF CHICKS INJECTED WITH 2 SEX HORMONES SEPARATELY AND IN COMBINATION

	Untreated			Hormone <i>B</i>		
	Number	<i>SX</i>	$\bar{x}$	Number	<i>SX</i>	$\bar{x}$
Untreated	3	240	80	12	1,200	100
Hormone <i>A</i>	12	1,440	120	6	672	112
	15	1,680		18	1,872	
Weighted mean	1,680/15 = 112			104		
Unweighted mean	(80 + 120)/2 = 100			106		

Preliminary Analysis of Variance, Ignoring Disproportion Among Subclass Numbers

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Hormone <i>A</i>	1	3,724	
Hormone <i>B</i>	1	524	
Individuals	29	23,519	811

Another startling characteristic of disproportionality in a 2-way table is the failure of the addition theorem for sums of squares. While this is not evident in table 11.18, it shows up in the analysis of the rat data of table 11.22, the two sums of squares, sex, and generation, adding to more than the total sum of squares. Clearly, no proper estimate of interaction is given in such a table. In fact, all estimates and tests of significance may be biased by the disproportion of subclass numbers, and the appropriate statistical methods are thereby complicated (27). Missing data in randomized blocks and latin square experiments constituted special cases of the more general one now to be explained.

In the problems of this section and the next two, there is required a preliminary analysis of variance in which the disproportion of subclass numbers is ignored. The analysis follows the pattern for proportional numbers in the preceding section. We are to learn how unbiased estimates and tests of significance may be attained for the main effects (means of rows and columns) and interactions. But it should be noted that there is no bias in the sums of squares for subclasses, either among or within them. In table 11.18, for example, the sums of squares among subclasses,

$$\frac{(240)^2}{3} + \frac{(1,440)^2}{12} + \frac{(1,200)^2}{12} + \frac{(672)^2}{6} - \frac{(3,552)^2}{33} = 4,940$$

and for individuals, 23,519, are correctly estimated in the manner of section 10.8. The terms "unequal subclass numbers" and "disproportionate subclass numbers" are not always clearly distinguished. The latter is often referred to as the *nonorthogonal* case: it occurs only in tables of multiple classification (more than one criterion).

A. INTERACTION NEGLIGIBLE

It will appear immediately that the presence or absence of interaction in the population determines the appropriate statistical method. Knowledge about interaction may come from theory or experience, or only from examination of the data. In the latter situation, it is desirable to have a convenient test, and this is available in $R \times 2$ tables (2 or more rows, 2 columns). The sum of squares for interaction is calculated directly from the subclass numbers and means (table 11.19), and this is tested for significance against the variance of individuals (table 11.18).

The calculations of table 11.19 require little explanation. Weights, W , are made up from the numbers of chicks, k , and these lead to a weighted sum of squares (1,216) of the mean differences. A correction term yields the interaction mean square which is little greater than sampling expectation. Therefore, the interaction in the population may be assumed negligible.

It turns out that the correction term, $(SWD)^2/SW$, is an unbiased

TABLE 11.19
INTERACTION NEGLIGIBLE IN A 2×2 TABLE WITH DISPROPORTIONATE SUBCLASS
NUMBERS. COMPUTATIONS FOLLOWING TABLE 11.18. CHICK HORMONE DATA

	Untreated		Hormone B		$\frac{k_1 k_2}{k_1 + k_2}$	$\bar{x}_2 - \bar{x}_1$	WD	WD^2
	k_1	$\bar{x}_1$	k_2	$\bar{x}_2$	$= W$	$= D$		
Untreated	3	80	12	100	2.4	20	48	960
Hormone A	12	120	6	112	4.0	-8	-32	256
					6.4		16	1,216

Sum of squares for interaction

$$= SWD^2 - (SWD)^2/SW = 1,216 - (16)^2/6.4 = 1,216 - 40 = 1,176$$

$$F = 1,176/811 = 1.45, d.f. = 1 \text{ and } 29, \text{ not significant}$$

Correction for disproportion

$$= (\text{sum of squares for hormone B}) - (SWD)^2/SW = 524 - 40 = 484$$

(see preliminary analysis, table 11.18)

Completed Analysis of Variance

Source of Variation	Degrees of Freedom	Mean Square
Hormone A	1	(3,724 - 484) 3,240
Hormone B	1	40
Interaction	1	1,176
Individuals	29	811

$$\text{For A, } F = 3,240/811 = 4.00; \text{ For B, } F = 40/811 = 0.05$$

estimate of the mean square between the column means (hormone B); hence, the preliminary value in table 11.18 must be reduced by a

$$\text{Correction for disproportion} = 524 - 40 = 484$$

The same correction for hormone A results in the completed analysis at the bottom of the table.

There is now available an unbiased estimate of the mean difference due to hormone B acting both with and without A: this weighted mean difference is:

$$SWD/SW = 16/6.4 = 2.5 \text{ mg.}$$

It is notable that this is not the difference between either the weighted or unweighted means of table 11.18. A similar estimate for the difference due to A may be calculated from the columns. In this table the weights happen to be the same as those in the rows so that

$$SWD/SW = [(2.4)(40) + (4.0)(12)]/6.4 = 22.5 \text{ mg.}$$

It is the significance of these weighted differences that is tested by the *F*-values in the table.

B. INTERACTION PRESENT

The Method of Unweighted Means (27)

This case is illustrated by table 11.20. From the preliminary analysis

TABLE 11.20

INTERACTION PRESENT. COMPUTATIONS FOR 2×2 TABLE WITH DISPROPORTIONATE SUBCLASS NUMBERS. THE DATA ARE DAYS TO DEATH OF 2 STRAINS OF MICE INOCULATED WITH 2 TYPHOID ORGANISMS (10)

Organism	Strain $\mathcal{Z}$		Strain <i>Ba</i>		$\frac{k_1 k_2}{k_1 + k_2}$	$\bar{x}_1 - \bar{x}_2$	<i>WD</i>	<i>WD</i> ²
	k_1	$\bar{x}_1$	k_2	$\bar{x}_2$	$= W$	$= D$		
9D	31	4.03	33	3.76	15.98	0.27	4.31	1.16
DSC 1	133	7.80	188	4.13	77.89	3.67	285.86	1,049.11
					93.87		290.17	1,050.27

Preliminary Analysis of Variance from the Original Data

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Subclasses	3	1,214.76	
Individuals	381	1,387.49	3.642

Interaction mean square

$$= SWD^2 - (SWD)^2/SW = 1,050.27 - (290.17)^2/93.87 = 153.3$$

$$F = 153.3/3.642 = 42.08, \text{ d.f.} = 1 \text{ and } 381$$

Harmonic mean of subclass numbers,

$$\frac{1}{k_0} = \frac{1}{4} \left(\frac{1}{31} + \frac{1}{33} + \frac{1}{133} + \frac{1}{188} \right) = 0.01885; k_0 = 53.05$$

of variance and the interaction mean square it is evident that interaction is present in the population—strain $\mathcal{Z}$ is relatively more resistant to DSC 1 than is *Ba*. The presence of interaction dictates the unweighted means as unbiased estimates of the strain, organism, and interaction effects in this 2×2 table.

Having discovered that the two typhoid organisms affect the strains of mice differentially, the investigator may not wish to go further. He will find that the means, 4.03, 3.76, and 4.13, differ only by the average amount expected in sampling from a single population ($F = 3.88/3.64$).

The outstanding result is the resistance of strain $\mathcal{Z}$ to DSC 1. It may be uninformative to average the rows or the columns. However, there are circumstances in which comparisons between the row means (and between the column means) are required. Hence, we proceed with the estimates and tests summarized in table 11.21.

TABLE 11.21
MEAN DAYS TO DEATH IN 4 GROUPS OF MICE, COPIED FROM TABLE 11.20.
CALCULATIONS WITH INTERACTION PRESENT

	$\mathcal{Z}$	Ba	Sum	Mean
9D	4.03	3.76	7.79	3.90
DSC 1	7.80	4.13	11.93	5.97
Sum	11.83	7.89	19.72	
Mean	5.92	3.94		4.93

$$\text{Mean squares: Organisms, } \frac{(11.93 - 7.79)^2}{4} = 4.28$$

$$\text{Strains, } \frac{(11.83 - 7.89)^2}{4} = 3.88$$

Completed Analysis of Variance

Source of Variation	Degrees of Freedom		Mean Square
Organisms	1	(4.28)(53.05)	227.1**
Strains	1	(3.88)(53.05)	205.8**
Interaction	1		153.3**
Individuals	381		3.642

If you have forgotten the easy way to calculate mean squares with a single degree of freedom, review section 10.5. Since these variances are based on means, they must be multiplied by some k to be comparable with those already computed; but because k is different for every cell of the table an average must be found. The appropriate one is the harmonic mean, $k_h = 53.05$ (table 11.20). The products of k_h by the mean squares first calculated complete the analysis of variance given in the table.

If one has ample reason to believe that interaction exists in the population, the major part of table 11.20 may be omitted—it was included to justify that hypothesis. If interaction is assumed to be present, its mean square may be calculated directly from the unweighted means:

$$\frac{[(7.80 + 3.76) - (4.03 + 4.13)]^2}{4} = 2.89$$

Then $(2.89)(53.05) = 153.3$, as before.

The three effects, strains, organisms and interaction, are all tested in the manner of section 11.8.

EXAMPLE 11.34—A good way to learn the effects of disproportionality and at the same time to gain experience in calculation is to complete some analysis incorrectly. For example, complete the analysis of the chick hormone data in table 11.19 on the assumption that interaction is not negligible in the population. Ans. Mean difference for *A*, 26; for *B*, 6 milligrams: mean square for *A*, 4,056; for *B*, 216.

EXAMPLE 11.35—Complete the analysis in table 11.20 on the assumption that interaction is negligible. Ans. Difference between organism means, 1.98; between strain means, 3.09 days: mean square for organisms, 896.97; for strains, 207.16.

11.11—Disproportionate subclass numbers. The $R \times 2$ table. This type of table is of rather common occurrence, ordinarily involving the two sexes. The computations, starting in the same way as those of the preceding section, diverge according to the presence or absence of interaction.

A. INTERACTION NEGLIGIBLE

The data in table 11.22 first brought to our attention some of the peculiarities of disproportionate subclass numbers. In the preliminary

TABLE 11.22

INTERACTION NEGLIGIBLE IN $R \times 2$ TABLE WITH DISPROPORTIONATE SUBCLASS NUMBERS. NUMBER AND MEAN GAIN IN WEIGHT (GRAMS) OF 149 WISTAR RATS DURING 1928-29. FOUR SUCCESSIVE GENERATIONS. GAINS DURING 6 WEEKS BEGINNING AT 28 DAYS OF AGE (4). 100 GRAMS SUBTRACTED FROM EACH GAIN

Generation	Male		Female		$\frac{k_1 k_2}{k_1 + k_2}$	$\bar{x}_1 - \bar{x}_2$	<i>WD</i>
	k_1	$\bar{x}_1$	k_2	$\bar{x}_2$	$= W$	$= D$	
1	21	76.952	27	9.518	11.8125	67.434	796.564
2	15	61.467	25	14.080	9.3750	47.387	444.253
3	12	55.667	23	8.522	7.8857	47.145	371.771
4	7	71.000	19	6.790	5.1154	64.210	328.460
					34.1886		1,941.048

Preliminary Analysis of Variance of Original Data

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Subclass means	7	119,141	
Sexes	1	114,287	
Generations	3	5,756	
Individuals	141		409

(Continued next page)

Interaction sum of squares: $*SWD^2 - (SWD)^2/SW = 113,385 - 110,202 = 3,183$
 Correction for disproportion = $114,287 - 110,202 = 4,085$
 Completed Analysis

Sexes	1	110,202	110,202
Generations	3 (5,756-4,085)	1,671	557
Interaction	3	3,183	1,061
Individuals	141		409

* SWD^2 is run up on the machine: $(67.434)(796.564) + \dots$

analysis of variance the two sums of squares, for sex and generation, add to more than the total for subclasses, emphasizing the nonorthogonal feature introduced by irregularities in the cell frequencies.

The interaction, though large, falls short of significance. Since there is no reason to believe that the difference in the sex rates of gain would change with generation, so far as the population is concerned, negligible interaction may be assumed.

The remaining computations proceed as in table 11.19.

This method is applicable to tables in which the data are missing entirely from a subclass. Zero may be entered instead of k and $\bar{x}$ with no disturbance of the computations. From the degrees of freedom for interaction deduct 1 for each cell with no entry.

The unbiased estimate of the sex difference in mean gain is

$$SWD/SW = 1,941.048/34.1886 = 56.77 \text{ grams}$$

That is the mean difference tested by $F = 110,202/409$.

Unbiased estimates of generation mean differences will be made available by the method of fitting constants to be explained in the next section. Very good approximations can be got by applying section 11.10-A, table 11.19. As an example, the weighted mean difference between generations 1 and 2 is calculated in table 11.23, and a test of significance is provided.

TABLE 11.23
 APPROXIMATE METHOD OF CALCULATING THE WEIGHTED MEAN DIFFERENCE BETWEEN GENERATIONS 1 AND 2. FOR EXACT METHOD, SEE SECTION 11.12-A

	Generation 1		Generation 2		<i>W</i>	<i>D</i>	<i>WD</i>
	<i>k</i> ₁	$\bar{x}_1$	<i>k</i> ₂	$\bar{x}_2$			
Male	21	76.952	15	61.467	8.75	15.485	135.49
Female	27	9.518	25	14.080	12.98	-4.562	-59.21
					21.73		76.28

Weighted mean difference = $SWD/SW = 3.51$ grams

Variance of mean difference = $409/21.73 = 18.82$ $s_d = 4.34$

$t = 3.51/4.34 = 0.81$, $d.f. = 141$

B. INTERACTION PRESENT
The Method of Weighted Squares of Means

For illustration we take the data in table 11.24 (21). The soils in the county were rated according to quality from high (I) to low (IV), and the farms classified according to rating and tenure. Among the tenants there seems to be a tendency to devote all suitable land to the production of corn.

The preliminary calculations proceed as before. While the interaction is not significant, it followed the same pattern in succeeding years and is

TABLE 11.24
ACRES IN CORN AMONG OWNERS AND TENANTS, AUDUBON COUNTY, IOWA, ON 4 SOIL GROUPS, 1934

Soil Group	Owners		Tenants		<i>W</i>	<i>D</i>	<i>WD</i>
	k_1	$\bar{x}_1$	k_2	$\bar{x}_2$			
I	36	32.7	67	55.2	23.418	22.5	526.90
II	31	36.0	60	53.4	20.440	17.4	355.66
III	58	30.1	87	46.8	34.800	16.7	581.16
IV	43	41.6	68	43.5	26.342	1.9	50.05
					105.000		1,513.77

From preliminary analysis, mean square for farms = 830
degrees of freedom = 442

Sum of squares for interaction = $SWD^2 - (SWD)^2/SW = 6,020.4$
Mean square = $6,020.4/3 = 2,006.8$

$F = 2,006.8/830 = 2.42$, $d.f. = 3$ and 442 , $F_{.05} = 2.62$

probably a population characteristic. We shall proceed on that assumption, computing the analysis of variance in table 11.25. Yates called this the method of *weighted squares of means* (27). It duplicates the weights in table 11.24, though a different method of calculation is convenient to get the weights for the columns: here, the reciprocals are put down individually, then added—*W* is the reciprocal of the sum. The weights are now applied to the squares of the unweighted means (WD^2) in order to reach unbiased sums of squares.

The factors, 4 and 16, the squares of the numbers of columns and of rows, are required because means instead of sums are used in the computations.

If one is sure there is interaction in the population, the preliminary computations of table 11.24 may be combined with those of 11.25 in a single table.

TABLE 11.25

COMPUTATION OF ANALYSIS OF VARIANCE IN $R \times 2$ TABLE WITH INTERACTION PRESENT.
WEIGHTED SQUARES OF MEANS. DATA FROM TABLE 11.24

Soil Group		Owners	Tenants	Sum of Reciprocals	Weights W	Un-weighted Means $\bar{x}$	$W\bar{x}$
I	k	36	67				
	$1/k$	0.027778	0.014925	0.042703	23.418		
	$\bar{x}$	32.7	55.2			43.95	1,029.22
II	k	31	60				
	$1/k$	0.032258	0.016667	0.048925	20.439		
	$\bar{x}$	36.0	53.4			44.70	913.62
III	k	58	87				
	$1/k$	0.017241	0.011494	0.028735	34.801		
	$\bar{x}$	30.1	46.8			38.45	1,338.10
IV	k	43	68				
	$1/k$	0.023256	0.014706	0.037962	26.342		
	$\bar{x}$	41.6	43.5			42.55	1,120.85
					105.000		4,401.79
Sum of reciprocals		0.100533 + 0.057792 = 0.158325 = sum of column, also.					
Weights, W		9.9470 + 17.3034 = 27.2504					
Unweighted means, $\bar{x}$		35.10	49.72			42.41	
$W\bar{x}$		349.14	+860.33				= 1,209.47

Weighted sums of squares:

Soils: $4[SW\bar{x}^2 - (SW\bar{x})^2/SW] = 4[185,215.15 - (4,401.79)^2/105.000] = 2,736.8$
 Tenancy: $16[SW\bar{x}^2 - (SW\bar{x})^2/SW] = 16[55,030.42 - (1,209.47)^2/27.2504] = 21,597.3$

Completed Analysis of Variance

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Soil groups	3	2,736.8	912
Tenancy	1	21,597.3	21,597
Interaction	3	6,020.4	2,007
Farms	442		830

On the assumption that interaction is present, the unweighted differences and means in the rows and columns are unbiased estimates. The weighted squares are used to calculate variance only.

EXAMPLE 11.36—Becker and Hall (2) determined the number of oocysts produced by rats of 5 strains during immunization with *Eimeria miyairii*. The unit of measurement is 10^6 oocysts.

Sex		Strain				
		Lambert	Lo	Hi	W.E.L.	Wistar (A)
Male	k	8	14	20	8	9
	$\bar{x}$	36.1	94.9	194.4	64.1	175.7
Female	k	7	14	21	10	8
	$\bar{x}$	31.9	68.6	187.3	89.2	148.4

Verify the completed analysis of variance quoted from the original article:

Sex	1	2,594.6	2,594.6
Strain	4	417,565.6	104,391.4
Interaction	4	8,805.3	2,201.3
Error	109	332,962.9	3,054.7

You will not be able to duplicate these numbers exactly because the means are reported to only 3 significant digits. Your results should approximate the first 3 figures in the mean squares, and no more are necessary for testing.

EXAMPLE 11.37—Another set of mice from Gowen's investigations (10) yielded the following data on days to death:

Strain		Organism		
		9D	11C	DSC 1
$\mathcal{Z}$	k	172	70	24
	$\bar{x}$	4.42	6.97	8.38
Ba	k	182	110	33
	$\bar{x}$	3.63	4.43	4.45

The variance of the individual days to death was 3.54. Compute the mean square for interaction, 84.9, and complete the analysis by the appropriate method of weighted squares. Ans. Mean square for organisms, 247; for strains, 494.

11.12—Disproportionate subclass numbers. The $R \times C$ table. Now we shall give the general methods of which those in the last two sections are special cases. Ordinarily the foregoing easier methods are adequate, but the more extensive tables appear often enough to call for the appropriate treatment.

The easy method of testing interaction is not available for tables with more than 2 rows and columns. This requires shrewder appraisal of the situation, because much time can be lost in calculation only to learn that the wrong method has been chosen. Examine the succession of differences in the several rows or columns: if they conform to a pattern, inter-

action is likely negligible. The size of the error mean square, together with the degrees of freedom available, helps in forming judgments.

In case of doubt, the method of unweighted means, section 11.10-B, (not hard to compute) may furnish the clue: this is especially true if the subclass numbers are not very different. While this method of unweighted means is not exact in the general table, though it was exact in the earlier section, it may convince you that no further analysis is necessary.

An illustration is found in table 11.26, abstracted as a representative sample from more extensive tabulations (6). The object of the investigation was to learn if season affected these gains. As explained before, the subclass numbers are ignored, the method of calculation being that of section 11.2 applied to the mean gains. All calculations are indicated in the table.

The completed analysis shows that the variation in weight-gains is only such as may be expected in sampling from a single normal population. The weighted mean of all the gains, 33.8 pounds per month, is the appropriate estimate in this undifferentiated sample, and $s = \sqrt{295} = 17.2$ pounds per month measures the variation.

TABLE 11.26
MONTHLY MEAN GAINS IN WEIGHT (POUNDS) OF HOLSTEIN HEIFERS AT THREE AGES

Month of Gain	Age in Months					
	13		18		24	
	<i>k</i>	$\bar{x}$	<i>k</i>	$\bar{x}$	<i>k</i>	$\bar{x}$
August	5	22.2	7	21.4	3	40.7
September	8	24.4	3	30.3	1	39.0
October	4	45.0	3	30.3	3	48.3
November	4	18.0	3	67.0	6	49.0

Mean square of individuals within subclasses = 931.8,
d.f. = 38 (from original data)

$$\frac{1}{k_0} = \frac{1}{12} \left(\frac{1}{5} + \frac{1}{8} + \dots + \frac{1}{3} + \frac{1}{6} \right) = 0.3168$$

Error mean square for comparison with mean squares calculated
from means = (931.8)(0.3168) = 295

Completed Analysis of Variance of Means

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Months	3	561	187
Ages	2	573	287
Interaction	6	1,185	198
Error	38		295

If some of the F -ratios had been near the levels of significance, one of the exact methods, A or B to follow, would be available: the approximate method would presumably afford an accurate enough test of interaction to warrant decision as to whether A or B should be used.

If the subclass numbers are quite different, especially if they seem to be characteristic of the population, another exploratory method may be available—the method of proportional subclass numbers (18) (19). The method is illustrated in table 11.27. In section 9.9 the observed subclass

TABLE 11.27
FARM ACRES IN CORN CLASSIFIED BY TENURE AND SOIL PRODUCTIVITY.
AUDUBON COUNTY, IOWA (21)

Soil Class		Owner		Renter		Mixed		Sk	SX
		Observed	Expected	Observed	Expected	Observed	Expected		
I	k	36	36.75	67	62.92	49	52.33	152	7,323
	$\bar{x}$	32.7		55.2		50.6			
	SX		1,202		3,473		2,648		
II	k	31	33.85	60	57.95	49	48.20	140	6,584
	$\bar{x}$	36.0		53.4		47.1			
	SX		1,219		3,095		2,270		
III	k	58	54.40	87	93.13	80	77.47	225	9,102
	$\bar{x}$	30.1		46.8		40.1			
	SX		1,637		4,358		3,107		
Sk		125		214		178		517	
SX			4,058		10,926		8,025		23,009

Analysis of Variance Using Expected Sums and Numbers

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Soils	2	6,635	3,318
Tenure	2	27,367	13,684
Interaction	4	883	221
Error (from original data)	507		830

numbers were shown to be independent, so that the proportional numbers may be taken to represent the distribution in the population. The product of the proportional number by the subclass mean gives a value of SX , and these two new statistics are used to calculate the analysis of variance in the manner of section 11.9.

There is no evidence of interaction in the population, while the tenure means are doubtless different. If it seems necessary to make a more critical test of the soil effects, the method A below is appropriate.

A. INTERACTION NEGLIGIBLE

The Method of Fitting Constants

This method leads to a test of the significance of interaction. If interaction turns out to be negligible, the accompanying estimates and other tests of significance are unbiased, and the analysis is complete. But if interaction becomes evident, the main effects and their tests are only

TABLE 11.28

THE METHOD OF FITTING CONSTANTS. DAYS TO DEATH IN 3 STRAINS OF MICE INOCULATED WITH 3 ISOLATIONS OF THE TYPHOID BACILLUS

Organism		Strain of Mice			<i>Sk</i>	<i>SX</i>	$\bar{x}_R$	$b + \bar{x}$
		<i>RI</i>	$\bar{Z}$	<i>Ba</i>				
9D	<i>k</i>	34	31	33	98	385		3.9490
	<i>k/Sk</i>	0.34694	0.31633	0.33673			3.9286	
11C	<i>k</i>	66	78	113	257	1,442		5.8600
	<i>k/Sk</i>	0.25681	0.30350	0.43969			5.6109	
DSC 1	<i>k</i>	107	133	188	428	2,523		6.1375
	<i>k/Sk</i>	0.25000	0.31075	0.43925			5.8949	
<i>Sk</i>		207	242	334	783			
<i>SX</i>		1,271	1,692	1,387		4,350		
<i>s</i>		0.4506	1.2240	-1.6746				$\bar{x} = 5.3155$

Preliminary Analysis of Variance

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Subclasses	8	1,785.58	
Organisms	2	309.47	
Strains	2	1,227.19	
Individuals	774		5.015

$$\begin{aligned}
 a_1: & [(34)(0.34694) + \dots + (107)(0.25000) - 207]a_1 \\
 & + [(34)(0.31633) + \dots + (107)(0.31075)]a_2 \\
 & + [(34)(0.33673) + \dots + (107)(0.43925)]a_3 \\
 & = (34)(3.9286) + \dots + (107)(5.8949) - 1,271
 \end{aligned}$$

(and so on to the last *a*)

$$\begin{aligned}
 a_2: & [(33)(0.34694) + \dots + (188)(0.2500)]a_1 \\
 & + [(33)(0.31633) + \dots + (188)(0.31075)]a_2 \\
 & + [(33)(0.33673) + \dots + (188)(0.43925) - 334]a_3 \\
 & = (33)(3.9286) + \dots (188)(5.8949) - 1,387
 \end{aligned}$$

approximate: then the method following this may be applied. It is desirable, therefore, to anticipate interaction if possible—from theoretical or experiential considerations, from examination of the data or from the foregoing analysis of unweighted means. Of course, the discovery of the presence or absence of interaction may be the objective of the experiment, so that this method of fitting constants is final. On the other hand, interaction may be obvious, in which case, go directly to the following method of weighted squares for estimates and tests of the main effects.

More of Gowen's data (10) are used for illustration, table 11.28. There was strong evidence of interaction in tables 11.20 and 11.21, and it is anticipated here: we shall apply both methods for the sake of contrast.

The table is made up from the original data containing the subclass numbers and sums of days to death. If the $R \times C$ table is rectangular instead of square, arrange it so that R is large and C small—the fewer the columns the easier the calculations. Each row of ratios should add to unity. The means, $\bar{x}_R$, are weighted means, SX/Sk .

The 7 constants to be fitted by regression methods are an overall mean, $\bar{x}$, deviations in each of the columns, a , and in each of the rows, b . These deviations are

$$a = \bar{x}_C - \bar{x} \quad \text{and} \quad b = \bar{x}_R - \bar{x},$$

which could have been calculated in section 11.3 and used to get the regression values, $\hat{X}$: it can be shown quite easily that

$$\hat{X} = a + b + \bar{x}$$

After the constants are computed, they will be used to make a table of regression values with zero interaction like 11.4. Note that, since the a 's and b 's are deviations from $\bar{x}$, therefore $Sa = Sb = 0$.

Corresponding to each column an a -equation is written as outlined at the bottom of the table. In a_1 the subclass numbers in column 1 are used over and over, each time with successive columns of ratios; and in the right member, with the column of means. In a_2 the numbers in column 2 are applied to the same pattern of ratios and means. The only fault in the pattern is the subtraction of Sk from one of the terms in each equation.

The coefficients of the a 's, easily calculated with a machine, are set down in the following set of simultaneous linear equations:

$$\begin{aligned} -151.505 a_1 + 64.036 a_2 + 87.468 a_3 &= -136.35 \\ 64.036 a_1 - 167.191 a_2 + 103.155 a_3 &= -348.54 \\ 87.468 a_1 + 103.155 a_2 - 190.624 a_3 &= 484.92 \end{aligned}$$

In verification of the computations, observe that the coefficients occur in pairs symmetrically arranged with respect to the principal diagonal, and that the sum of the coefficients in each of the 3 columns, as well as the sum of the right hand members, is zero (except for rounding errors).

Since $Sa = 0$, the first step is to substitute $a_3 = -a_1 - a_2$ in each

equation. This is easily done: simply subtract 87.468 from each of the preceding coefficients in the first equation, and 103.155 from each in the second, the results being

$$\begin{aligned} -238.973 a_1 - 23.432 a_2 &= -136.35 \\ -39.119 a_1 - 270.346 a_2 &= -348.54 \end{aligned}$$

The third equation is merely the sum of these two and may be discarded. The solution is

$$\begin{aligned} a_1 &= 0.4506 \\ a_2 &= 1.2240 \\ a_3 &= -1.6746 \end{aligned}$$

and, therefore,

Verify by substitution in one or more of the original equations. It is convenient to enter the a 's in the last row of the table.

A quantity, $b + \bar{x}$, in each row is now calculated from the a 's together with $\bar{x}_R$ and the ratios in the same row. As an example, in row 1,

$$\begin{aligned} b_1 + \bar{x} &= 3.9286 - (0.34694)(0.4506) - \dots - (0.33673)(-1.6746) \\ &= 3.9490 \end{aligned}$$

Similarly,

$$\begin{aligned} b_2 + \bar{x} &= 5.8600 \\ b_3 + \bar{x} &= 6.1375 \end{aligned}$$

Adding,

$$\begin{aligned} 3\bar{x} &= 15.9465 \text{ (because } Sb = 0) \\ \bar{x} &= 5.3155 \end{aligned}$$

Finally,

$$\begin{aligned} b_1 &= -1.3665 \\ b_2 &= 0.5445 \\ b_3 &= 0.8220 \end{aligned}$$

The "reduction in sum of squares due to fitting constants" is now calculated from the constants and the original SX in the borders of the table by this process:

$$\begin{aligned} (0.4506)(1,271) + (1.2240)(1,692) + \dots + (6.1375)(2,523) - (4,350)^2/783 \\ = 1,609.78 \end{aligned}$$

Deducting this from the sum of squares for subclasses, the interaction sum of squares is attained,

$$1,785.58 - 1,609.78 = 175.80$$

Since there are 4 degrees of freedom for interaction, we have

$$\begin{aligned} \text{Mean square for interaction} &= 175.80/4 = 43.95 \\ F &= 43.95/5.015 = 8.76, df. = 4 \text{ and } 774, F_{.01} = 3.35 \end{aligned}$$

With rejection of the hypothesis of zero interaction, we are warned that further computations by this method yield somewhat biased estimates. The mean square for interaction is correctly estimated and tested, but we must turn to the method of weighted squares, following this, to get further information. However, as illustration of the remaining procedure if interaction were negligible, we complete the calculations.

With the constants now fitted, the regression values, $\hat{X} = a + b + \bar{x}$, are calculated in table 11.29. If there were no interaction, the means for

TABLE 11.29
REGRESSION VALUES FROM FITTED CONSTANTS. INTERACTION ASSUMED NEGLIGIBLE.
MOUSE DATA OF TABLE 11.27

Organism	Constants	<i>RI</i>	$\bar{Z}$	<i>Ba</i>	Mean
		$a_1 = 0.4506$	$a_2 = 1.2240$	$a_3 = -1.6746$	
9D	$b_1 = -1.3665$	4.3996	5.1730	2.2744	3.9490
11C	$b_2 = 0.5445$	6.3106	7.0840	4.1854	5.8600
DSC 1	$b_3 = 0.8220$	6.5881	7.3615	4.4629	6.1375
Mean		5.7661	6.5395	3.6409	

Completed Analysis

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Organisms	2	309.47 - (-73.12) = 382.59	191.30
Strains	2	1,227.19 - (-73.12) = 1,300.31	650.16
Interaction	4	175.80	43.95
Mice	774		5.015

strains and organisms would be the correct estimates of days to death. Also, the analysis of variance would be completed by a correction to be deducted from the sums of squares for organism and strain in the preliminary analysis:

$$\begin{aligned}\text{Correction} &= (\text{total for organism and strain}) - (\text{reduction due to constants}) \\ &= 1,536.66 - 1,609.78 = -73.12\end{aligned}$$

The completed analysis is entered at the bottom of the table.

The nature of the reduction in sum of squares due to fitting constants will be clear if you go through this process: calculate the sum of squares for the 9 subclass means in table 11.29, having no interaction, and compare it with the corresponding sum in table 11.28. The difference is the interaction sum of squares, but rounding errors have crept into the fourth significant figure.

B. INTERACTION PRESENT

The Method of Weighted Squares of Means

This method gives unbiased estimates and tests of the main effects if interaction is present. It furnishes no information about interaction in the general $R \times C$ table, though it did give this information in the special cases where one or both of R and C were 2. In the 2×2 table, the methods of unweighted means and of weighted squares are equivalent.

Since we now have evidence of what was anticipated about the mouse data, that interaction is present, we turn to the present method to complete the examination correctly. Table 11.30 is a simple extension of table 11.25, and requires no further explanation.

Before embarking on this computation, ask yourself this question: since the strains do not respond in the same way to the organisms, is anything to be gained by estimates and tests of the overall averages? On the

TABLE 11.30
COMPUTATIONS IN $R \times C$ TABLE WITH DISPROPORTIONATE SUBCLASS NUMBERS.
INTERACTION PRESENT
Weighted squares of means. Mouse data from table 11.28.

Organism		Strain			Sum of Reciprocal	Weight W	Un-weighted Means $\bar{x}$	$W\bar{x}$
		RI	Z	Ba				
9D	k	34	31	33	0.0919	10.873	3.9300	42.731
	$1/k$	0.0294*	0.0322	0.0303				
	$\bar{x}$	4.0000	4.0323	3.7576				
11C	k	66	78	113	0.0368	27.158	5.8488	158.842
	$1/k$	0.0152	0.0128	0.0088				
	$\bar{x}$	6.4545	6.7821	4.3097				
DSC 1	k	107	133	188	0.0221	45.078	6.1861	278.857
	$1/k$	0.0093	0.0075	0.0053				
	$\bar{x}$	6.6262	7.8045	4.1277				
						83.109		480.430
Sum of $1/k$		0.0539 +	0.0525 +	0.0444 =	0.1508 = sum of column, also.			
Weights, W		18.550 +	19.012 +	22.486 =	60.048			
Unw'ted $\bar{x}$		5.6936	6.2063	4.0650				
$W\bar{x}$		105.616 +	117.994 +	91.406 =	315.016			

Weighted sums of squares:

Organisms, $9[SW\bar{x}^2 - (SW\bar{x})^2/SW] = 403.11$

Strains, do. = 473.40

Completed Analysis of Variance

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Organisms	2	403.11	201.6
Strains	2	473.40	236.7
Interaction (table 11.28)	4		43.95
Mice	774		5.015

* The computations were made with 6 decimal places so rounding discrepancies will be found if only 4 are used.

assumption that the border means do represent some specifiable populations, we proceed with the illustration. Since interaction is present, we take the unweighted means as estimates, and arrive at a test of the significance of differences among them.

It is worth while to observe the effect of interaction by comparing table 11.30, in which it was properly taken into account, with 11.29 where it was ignored.

In closing these sections on disproportionate subclass numbers, it should be observed that use of the appropriate method of analysis ordinarily makes the efficiency of such experiments (that is, the amount of information per individual) close to 100%. Attention has been drawn to contrasts between the results of appropriate and inappropriate methods applied to the same data: clearly, much information can be lost by use of the wrong method. Other things being equal, data should be collected in equal or proportional numbers because the information per individual is maximized and the computations are easier. No statistical method can compensate for the loss of data—it is only the remaining information that can be extracted, and that at some loss of efficiency.

11.13—Experiments repeated in several places or years. Agricultural experiments are often repeated in various localities and in successive seasons in order to learn about various environmental effects on the varieties or treatments being studied. Each experiment is usually of some replicated design from which the local responses may be evaluated in the manner already described. The object of the repetition is to determine if one can make recommendations which are widely applicable.

Such groups of experiments pose a variety of problems for whose solution there is no simple set of rules. The difficulties are discussed by Yates and Cochran (30) who illustrated methods suitable for particular sets of data.

Some of the principles are applied to the data in table 11.31, selected from the returns from the 1943 cooperative seed treatment tests on Mukden soybeans (16). Four treated lots of 100 seeds, and 1 untreated, were planted in 5 randomized blocks at each participating station. The total numbers of emerging plants (from 500 seeds) are shown for 5 locations.

From these data, supplemented by the local analyses of variance at the bottom of the table, it is evident that conditions affecting emergence varied greatly from place to place. Only in some of the stations were the treatments effective. Even the error mean squares differed notably, a circumstance not unusual in this kind of investigation. Our problem is to discover any general features of the data that may be useful.

The first thing to learn is whether the mean squares for experimental error may be estimates of some common variance: if so, they can be combined into a single more reliable estimate. The method is that of section 10.13 with $n = 5$ and $k - 1 = 16$. Corrected chi-square turns out to

TABLE 11.31
NUMBER OF EMERGING PLANTS (500 SEEDS) IN 5 PLOTS. COOPERATIVE SEED
TREATMENT TRIALS WITH MUKDEN SOYBEANS. 1943

Location	Untreated	Arasan	Spergon	Semesan, Jr.	Fermate	Total
Michigan	360	356	362	350	373	1,801
Minnesota	302	354	349	332	332	1,669
Wisconsin	408	407	391	391	409	2,006
Virginia	244	267	293	235	278	1,317
Rhode Island	373	387	406	394	375	1,935
Total	1,687	1,771	1,801	1,702	1,767	8,728

Mean Squares From Original Analyses of Variance

Source of Variation	Degrees of Freedom	Location				
		Michigan	Minnesota	Wisconsin	Virginia	Rhode Island
Treatment	4	14.44	82.84*	17.44	114.26*	37.50
Block	4	185.14	54.64	5.64	70.76	4.80
Error	16	42.29	26.67	30.64	26.34	13.05

be 5.22, $d.f. = 4$, indicating little more than average sampling variation from a common σ^2 ; hence, we shall pool the 5 error mean squares.

Among all the states taking part in this experiment, the error mean squares varied significantly; five with homogeneous variance were picked out to illustrate what can be done with a common estimate. If error is heterogeneous, one special method which can be used will be indicated later.

The 5 local experiments can be combined into a single analysis of variance as in table 11.32. For replications and error, the degrees of freedom as well as the sums of squares from the several stations are pooled in the manner of chapters 4 and 10. On the other hand, since each treat-

TABLE 11.32
ANALYSIS OF VARIANCE OF 5 RANDOMIZED BLOCKS EXPERIMENTS. THE SOYBEAN DATA

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Treatments	4	380.29	95.07
Location	4	11,852.61	2,963.15
Interaction	16	685.63	42.85
Replications	20	1,283.92	
Experimental error	80	2,223.68	27.80

For interaction: $F = 42.85/27.80 = 1.54$, $F_{.05} = 1.77$

For treatment $F = 95.07/42.85 = 2.22$, $F_{.05} = 3.01$ (see text)

ment is repeated at all places, the totals in the 2-way table, 11.31, yield sums of squares for location, for treatment, and for interaction, the calculation being that of steps 3, 5, 6, and 7, section 11.8. Note that each cell contains the sum of 5 plot records.

Since the pooled error mean square is an average of homogeneous local errors, it may be used to test the interaction; and also the treatment effect *in the particular set of places chosen*. Thus, in this experiment the value of F for treatments is $95.07/27.80 = 3.42$, $P = 0.02$. But ordinarily this test is not of much interest. What we wish to know is whether a treatment is likely to be superior not only here but elsewhere in the region. Evidence on this depends upon the manner in which the places for the experiments were selected. If they can be considered a random sample of all places in the specified region, then the test for treatments is that given in the table—a component must be included in the error mean square (the treatment-place interaction) to allow for the fact that treatment differences may vary from place to place over the sampled region. The result of the test indicates that no general recommendation can be made with reasonable confidence.

In this example, the stations are certainly not randomly chosen, nor were they in the entire cooperative seed treatment trial, though the whole group of stations might be considered representative of some region. The requirement of randomness is apparently never met in this kind of work, so that statements about probability must be considered inexact.

Since the treatment and interaction mean squares are both larger than experimental error, it is appropriate to look for sources of the additional variation. One naturally asks if emergence in the treated plots as a group differs from that in the untreated. As before (section 11.8), the data are condensed into the form:

	Michigan	Minnesota	Wisconsin	Virginia	Rhode Island
Untreated	360	302	408	244	373
Treated	1,441	1,367	1,598	1,073	1,562

The results in table 11.33 show that substantial portions of the sums of squares of both treatment and interaction are due to the more numerous emerging plants from the treated seeds, but the remaining mean squares are still larger than experimental error.

From further examination of table 11.31 it appears that Semesan Jr. tends to be less effective than the other treatments except in Wisconsin and Rhode Island. Another 2×5 table with emerging plants from Semesan Jr. treated seeds in one line and totals from the other 3 treatments in the second, with another set of computations like those in section 11.8, results in the sums of squares for "Semesan Jr. vs. others" in table 11.33. Deduction of these reduces the remaining mean squares to amounts easily attributed to experimental variation.

TABLE 11.33
ANALYSIS OF THE SOYBEAN SEED TREATMENT DATA TO ISOLATE SOURCES OF VARIATION

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Treatments:	(4)	(380.29)	
Untreated vs. treated	1	171.70	171.70
Semesan Jr. vs. others	1	180.88	180.88
Differences among remaining 3	2	27.71	13.86
Interaction:	(16)	(685.63)	
Untreated vs. treated	4	235.78	58.94
Semesan Jr. vs. others	4	189.84	47.46
Differences among remaining 3	8	260.01	32.50
Experimental error	80		27.80

It is now plain that Arasan, Spergon, and Fermate have substantially the same effects in the several places, but that untreated plots and those planted with Semesan Jr. treated seeds introduce variation beyond that measured by experimental error.

It remains to be said that the F -test for a comparison like untreated vs. treated, involving a single degree of freedom,

$$F = 171.70/58.94 = 2.91, d.f. = 1 \text{ and } 4,$$

is little affected by the heterogeneity of experimental error usually encountered in agricultural experiments. This kind of comparison is more easily tested by t : write down the 5 pairs of means (for untreated and treated seeds, respectively) in parallel columns, then proceed as in table 2.2. The resulting t will be the same as $\sqrt{F} = \sqrt{2.91} = 1.71$, $d.f. = 4$.

11.14—Three criteria of classification. With increasing knowledge about experimental material, the researcher may be able to extend the bases of his judgments about outcome. Not only sex and treatment may affect weight gains but also strain or generation, so that further classifications are required in the analysis of variance. Another source of multiple classification is increasing complexity of treatment combinations. Both sources are involved in an experiment described by Winsor and Clarke (25), the object being to investigate the catch of various groups of plankton by 2 nets hauled horizontally, one 2 meters below the other. Ten hauls were made with the pair of nets at depths 29 and 31 meters respectively. It was observed that the standard deviations in the various groups were roughly proportional to the means, so that it seemed appropriate to examine the logarithms of the catch numbers (section 16.7). These are entered as the records in table 11.34.

Each item in the table is classified by group, by haul and by position, but the total sum of squares is calculated, as usual, from the 100 entries, ignoring classification.

$$\text{Correction: } (242.36)^2/100 = 587.3837$$

$$\text{Total: } (2.93)^2 + \dots + (1.45)^2 - 587.3837 = 31.3699$$

This total is entered at the bottom of the table of analysis of variance.

The observations are now combined in three 2-way tables, 11.34. As examples: in *A*, haul 1, group I, 5.54 is the sum of the catches in the 2 nets, $2.93 + 2.61$; in *C*, group I, upper position, $28.69 = 2.93 + 3.05 + \dots + 3.03$. In each of these tables the calculations are like steps 3, 5, 6, and 7, section 11.8. Using *B* as illustration:

$$3. \text{ Subclasses: } \frac{(12.94)^2 + (11.60)^2 + \dots + (10.76)^2}{5} - 587.3837 = 4.1371$$

$$5. \text{ Hauls: } \frac{(24.54)^2 + \dots + (22.83)^2}{10} - 587.3837 = 2.9969$$

EXAMPLE 11.38—In Minnesota during 1931, five varieties of barley were planted at 6 places, the yields (bushels per acre) being (12):

Locality	Block	Variety					
		Manchuria	Glabron	Trebi	Peatland	Wisconsin	Total
University	1	29	45	41	38	44	197
	2	25	39	40	30	37	171
	3	27	46	37	30	36	176
	Sum	81	130	118	98	117	544
Waseca	1	48	55	64	41	61	269
	2	52	53	64	52	58	279
	3	47	57	64	53	57	278
	Sum	147	165	192	146	176	826
Morris	1	24	28	42	25	32	151
	2	25	26	42	30	26	149
	3	34	33	47	35	31	180
	Sum	83	87	131	90	89	480
Crookston	1	38	40	53	40	45	216
	2	38	38	40	37	59	212
	3	44	37	47	47	46	221
	Sum	120	115	140	124	150	649
Grand Rapids	1	34	21	38	36	32	161
	2	27	30	29	39	36	161
	3	38	36	22	29	36	161
	Sum	99	87	89	104	104	483
Duluth	1	29	31	37	35	30	162
	2	27	33	33	28	30	151
	3	32	25	32	33	35	157
	Sum	88	89	102	96	95	470
Total		618	673	772	658	731	3452

This example is extracted from more extensive data examined by Yates and Cochran (30). Can you isolate some of the sources of variation?

$$6. \text{ Positions: } \frac{(124.50)^2 + (117.86)^2}{50} - 587.3837 = 0.4409$$

$$7. H \times P: 4.1371 - (2.9969 + 0.4409) = 0.6993$$

The sum of squares for *second order* (or *three factor*) *interaction*, group $\times$ haul $\times$ position, is the difference between the *total* and the accumulated sums of squares of the first 6 effects in the table. The corresponding

TABLE 11.34
LOGARITHMS OF CATCHES OF 5 GROUPS OF PLANKTON IN 2 NETS HAULED 10 TIMES

Haul	Position	Group					Total
		I	II	III	IV	V	
1	Upper	2.93	2.60	2.69	3.04	1.68	12.94
	Lower	2.61	2.34	2.41	3.06	1.18	11.60
2	Upper	3.05	2.42	2.93	2.91	1.94	13.25
	Lower	2.65	2.13	2.58	2.65	1.38	11.39
3	Upper	3.15	2.34	3.27	2.90	2.17	13.83
	Lower	3.02	2.19	3.22	2.88	2.01	13.32
4	Upper	2.54	1.65	2.59	2.44	1.97	11.19
	Lower	2.56	1.82	2.74	2.69	1.76	11.57
5	Upper	2.65	1.90	2.26	2.85	1.68	11.34
	Lower	2.30	1.61	2.15	2.75	1.15	9.96
6	Upper	2.85	2.08	2.98	2.38	1.81	12.10
	Lower	3.16	2.12	3.12	2.69	1.97	13.06
7	Upper	3.07	2.04	3.76	2.96	1.81	13.64
	Lower	3.08	2.03	3.64	2.47	1.77	12.99
8	Upper	2.44	1.58	3.03	2.68	1.74	11.47
	Lower	2.58	1.65	3.04	2.61	1.54	11.42
9	Upper	2.98	1.86	3.28	2.60	1.95	12.67
	Lower	2.92	1.84	2.91	2.54	1.58	11.79
10	Upper	3.03	2.06	2.60	2.64	1.74	12.07
	Lower	2.79	1.61	2.52	2.39	1.45	10.76
Total		56.36	39.87	57.72	54.13	34.28	242.36

Analysis of Variance

Source of Variation	Degrees of Freedom	Sum of Square	Mean Square
Main effects:			
Groups (G)	4	22.7583	5.6896
Hauls (H)	9	2.9969	0.3330
Positions (P)	1	0.4409	0.4409
First order interactions:			
G $\times$ H	36	3.9938	0.1109
G $\times$ P	4	0.1251	0.0313
H $\times$ P	9	0.6993	0.0777
Second order interaction:			
G $\times$ H $\times$ P	36	0.3556	0.0099
Total	99	31.3699	

TABLE 11.34—(Continued)
Three 2-Way Tables, Summary of the Plankton Data

A	Group					Total
Haul	I	II	III	IV	V	
1	5.54	4.94	5.10	6.10	2.86	24.54
2	5.70	4.55	5.51	5.56	3.32	24.64
3	6.17	4.53	6.49	5.78	4.18	27.15
4	5.10	3.47	5.33	5.13	3.73	22.76
5	4.95	3.51	4.41	5.60	2.83	21.30
6	6.01	4.20	6.10	5.07	3.78	25.16
7	6.15	4.07	7.40	5.43	3.58	26.63
8	5.02	3.23	6.07	5.29	3.28	22.89
9	5.90	3.70	6.19	5.14	3.53	24.46
10	5.82	3.67	5.12	5.03	3.19	22.83
Total	56.36	39.87	57.72	54.13	34.28	242.36

B	Haul										Total
Position	1	2	3	4	5	6	7	8	9	10	
Upper	12.94	13.25	13.83	11.19	11.34	12.10	13.64	11.47	12.67	12.07	124.50
Lower	11.60	11.39	13.32	11.57	9.96	13.06	12.99	11.42	11.79	10.76	117.86
Total	24.54	24.64	27.15	22.76	21.30	25.16	26.63	22.89	24.46	22.83	242.36

C	Group					Total
Position	I	II	III	IV	V	
Upper	28.69	20.53	29.39	27.40	18.49	124.50
Lower	27.67	19.34	28.33	26.73	15.79	117.86
Total	56.36	39.87	57.72	54.13	34.28	242.36

degrees of freedom are the product of those for the 3 main effects, $(4)(9)(1) = 36$.

The three factor interaction indicates:

1. The difference between the haul $\times$ group interactions in the 2 positions; or

2. The variation of the haul $\times$ position interactions in the 5 groups; or

3. The variation of the group $\times$ position interactions in the 10 hauls.

Components of variance in a 3-way table are as follows:

Groups	$\sigma^2 + 10\sigma_{GP}^2 + 2\sigma_{GH}^2 + 20\sigma_G^2$
Hauls	$\sigma^2 + 5\sigma_{HP}^2 + 2\sigma_{GH}^2 + 10\sigma_H^2$
Position	$\sigma^2 + 5\sigma_{HP}^2 + 10\sigma_{GP}^2 + 50\sigma_P^2$
$G \times H$	$\sigma^2 + 2\sigma_{GH}^2$
$G \times P$	$\sigma^2 + 10\sigma_{GP}^2$
$H \times P$	$\sigma^2 + 5\sigma_{HP}^2$
$G \times H \times P$	σ^2

The coefficients of the sigmas are the numbers of items making up the corresponding cell and border totals in table 11.34. In *B*, for instance, each cell total is the sum of 5 items; hence, the coefficient of σ_{HP}^2 is 5. Similarly, the coefficient of σ_H^2 is 10 because each haul total comprises 10 observations.

By equating the sample estimates in the analysis of variance to corresponding component sums it is easy to compute $s^2 = 0.0099$, $s_{HP}^2 = 0.0136$, $s_{GP}^2 = 0.0021$, $s_{GH}^2 = 0.0505$, $s_P^2 = 0.0068$, $s_H^2 = 0.0154$, and $s_G^2 = 0.2779$.

The component analysis is worked out on the following assumptions: (i) The variation of any *X* about *m* has 7 sources, one common to all members of the same group, one to all members of the same haul, etc. For a description of this common variation, review the sampling of table 10.9. (ii) The various sources of variation are independent and their effects additive.

Second order interaction often does not seem to be due to any real biological relationship but only to random deviations such as those measured by discrepance in the randomized blocks experiment (section 11.3). This second order discrepance is, therefore, commonly taken as an estimate of experimental error; that is, of σ^2 in a normally distributed population with $m = 0$. If one wishes to estimate and test second order interaction, he should provide two or more experimental units instead of one in each cell of the 3-way table, as was done in the 2-way table 11.14.

On the assumption that s^2 measures only random variation, it becomes interesting to use it for testing the significance of other effects. For testing the hypothesis that any first order interaction is zero, the *F* ratio is clearly indicated by the components of variance. Thus, for testing $\sigma_{GP}^2 = 0$,

$$F = \frac{s^2 + 10s_{GP}^2}{s^2} = \frac{0.0313}{0.0099} = 3.16, \quad P = 0.03$$

While all the two-factor interactions are significant, examination of table 11.34 fails to reveal any very striking contrasts. The most notable seems to be the large difference between the catches of group V in the two nets.

The method of testing a null hypothesis about a main effect depends on the assumptions made. In testing $\sigma_P^2 = 0$, for example, one might assume that the group and haul means are unchanging parameters; that is, that the test is to apply to this particular experiment only. Then,

$$F = 0.4409/0.0099 = 44.54, \quad d.f. = 1 \text{ and } 36$$

But under this assumption the test is not very interesting. It could well be confined to the specific groups enumerated, but not to the same hauls. A more reasonable assumption is that the hauls constitute a random sample from some population for which $\sigma_P^2 = 0$ may be tested, the groups being unvarying as before. In that case, the component for the position

differences with haul, $5s_{HP}^2$, must be added to the estimate of error, and $F = 0.4409/0.0777 = 5.67$, $df. = 1$ and 9 , $P = 0.05$

Concerning the position and position $\times$ group effects, Winsor and Clarke said, "The interpretation to be put on these figures is not entirely clear to us. Their statistical significance is well enough established. . . . We hesitate, however, to assert that they give a true picture of the distribution in depth of the plankton organisms." Other data led them to suspect that the results in this experiment may have been due to a larger net hauled between the two here reported.

Among the components of variance, two stand out; $s_{GH}^2 = 0.0505$ and $s_G^2 = 0.2779$. The latter is due to the differences in abundance of the groups of organisms. The former seems to reflect changing differences among the groups during the 4 hours occupied by the experiment.

EXAMPLE 11.39—Here is a 3-way table for practice: two observations in each cell of the table.

Species	Sex	Treatments			
		1	2	3	4
A	Male	8, 10	8, 12	7, 11	7, 9
	Female	3, 7	6, 8	3, 3	1, 1
B	Male	5, 9	7, 7	9, 11	11, 13
	Female	3, 3	4, 8	5, 9	6, 10
C	Male	7, 9	6, 14	12, 16	14, 18
	Female	4, 4	4, 6	10, 12	12, 12

Analysis of Variance

Source of Variation	Degrees of Freedom	Mean Square
Treatments	3	30
Species	2	52
Sex	1	192
Treatment-species	6	30
Treatment-sex	3	2
Species-sex	2	4
Treatment-species-sex	6	2
Individuals	24	5
Total	47	

11.15—The split plot experiment. It is often desirable to get rather precise information on one set of treatments as well as on the interaction of this set with a second, but to forego such precision on the second set. For example, 3 sources of a vitamin might be compared by trying them on 3 males of the same litter, replicating the experiment on 20 litters. This

would be a randomized blocks design with 38 *d.f.* for error. Superimposed on this could be some experiment with the litters as units: four types of housing could be tried, one litter to each type, thus allowing 5 replications. The main treatments (housings) would not be compared so accurately as the subtreatments (sources of vitamin) for two reasons; less replication is provided, and litter differences are included in the error for evaluating the housing effects. Nevertheless, some information about housing may be got at little extra expense, and any interaction between housing and vitamin will be accurately evaluated.

TABLE 11.35
YIELDS OF 3 VARIETIES OF ALFALFA (TONS PER ACRE) IN 1944 FOLLOWING 4 DATES OF FINAL CUTTING IN 1943

Variety	Date	Block					
		1	2	3	4	5	6
Ladak	A	2.17	1.88	1.62	2.34	1.58	1.66
	B	1.58	1.26	1.22	1.59	1.25	0.94
	C	2.29	1.60	1.67	1.91	1.39	1.12
	D	2.23	2.01	1.82	2.10	1.66	1.10
Cossack		8.27	6.75	6.33	7.94	5.88	4.82
	A	2.33	2.01	1.70	1.78	1.42	1.35
	B	1.38	1.30	1.85	1.09	1.13	1.06
	C	1.86	1.70	1.81	1.54	1.67	0.88
Ranger	D	2.27	1.81	2.01	1.40	1.31	1.06
		7.84	6.82	7.37	5.81	5.53	4.35
	A	1.75	1.95	2.13	1.78	1.31	1.30
	B	1.52	1.47	1.80	1.37	1.01	1.31
	C	1.55	1.61	1.82	1.56	1.23	1.13
	D	1.56	1.72	1.99	1.55	1.51	1.33
		6.38	6.75	7.74	6.26	5.06	5.07
Total		22.49	20.32	21.44	20.01	16.47	14.24

Variety	Date of Cutting				Total
	A	B	C	D	
Ladak	11.25	7.84	9.98	10.92	39.99
Cossack	10.59	7.81	9.46	9.86	37.72
Ranger	10.22	8.48	8.90	9.66	37.26
Total	32.06	24.13	28.34	30.44	114.97
Mean (tons per acre)	1.78	1.34	1.57	1.69	

The data in table 11.35 are from a split plot field trial on alfalfa. Three varieties were planted on the main plots, and these were replicated 6 times (24). As shown in figure 11.3, each of the 18 main plots was split into 4 subplots on which there were 4 dates of final cutting: *A*, no third

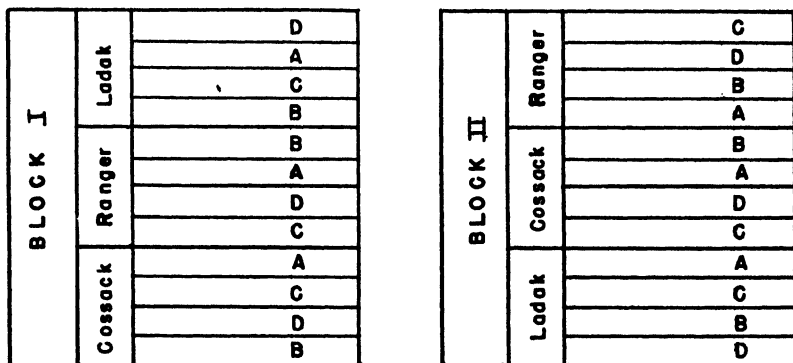


FIG. 11.3—First 2 blocks of split plot experiment on alfalfa, illustrating random arrangement of main and sub-plots.

cutting, the second harvest on all plots being on July 27; *B*, third cutting on September 1, 1943; *C*, September 20; and *D*, October 7. Not only must the varieties be randomized on the main plots, but the dates must be assigned at random to the subplots of each main plot.

The analysis of variance in table 11.36 reflects the limitations imposed by the split plot design. The main plot analysis is that of a randomized blocks experiment with 3 varieties replicated in 6 blocks. In the subplot analysis, the 4 dates of cutting are replicated on each of the 18 main

TABLE 11.36
ANALYSIS OF VARIANCE OF SPLIT PLOT EXPERIMENT ON ALFALFA

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Main plots:			
Varieties	2	0.1781	0.0890
Blocks	5	4.1499	0.8300
Main plot error	10	1.3622	0.1362
Subplots:			
Dates of cutting	3	1.9625	0.6542
Date $\times$ variety	6	0.2105	0.0351
Subplot error	45	1.2586	0.0280

For varieties: $F = 0.0890/0.1362 = 0.65$

For dates: $F = 0.6542/0.0280 = 23.36$

Date $\times$ variety: $F = 0.0351/0.0280 = 1.25$

plots. The error for subplots is usually smaller than that for main plots because the smaller plots lie closer together with less hazard of soil differences.

A comparison of this analysis with that of the 3-way table of the plankton sampling shows: (i) that the interaction between variety and block is the error for testing main treatments; and (ii) that there is no real second order interaction, but instead a pooling of the date-block discrepancies (with $(4 - 1)(6 - 1) = 15$ *df.*) within the 3 varieties, leading to the second error (with $3 \times 15 = 45$ *df.*) for testing subtreatments and their interaction with main treatments.

The computations are as follows:

1. Correction: $C = (114.9)^2/72 = 183.5847$

2. Total: $(2.17)^2 + \dots + (1.33)^2 - C = 9.1218$

3. Main plot: $\frac{(8.27)^2 + \dots + (5.07)^2}{4} - C = 5.6902$

4. Variety: $\frac{(39.99)^2 + \dots + (37.26)^2}{24} - C = 0.1781$

5. Block: $\frac{(22.49)^2 + \dots + (14.24)^2}{12} - C = 4.1499$

6. Main plot error: $5.6902 - (0.1781 + 4.1499) = 1.3622$

7. Subclasses in variety-date table:

$$\frac{(11.25)^2 + \dots + (9.66)^2}{6} - C = 2.3511$$

8. Date: $\frac{(32.06)^2 + \dots + (30.44)^2}{18} - C = 1.9625$

9. Variety $\times$ date: $2.3511 - (0.1781 + 1.9625) = 0.2105$

10. Subplot error: $9.1218 - (5.6902 + 1.9625 + 0.2105) = 1.2586$

The significant differences among dates of cutting were not unexpected, nor were the smaller yields following *B* and *C*. The last harvest should be either early enough to allow renewed growth and restoration of the consequent depletion of root reserves, or so late that no growth and depletion will ensue. The surprising features of the experiment were two: the yield following *C* being greater than *B*, since late September is usually considered a poor time to cut alfalfa in Iowa; and the absence of interaction between date and variety—Ladak is slow to renew growth after cutting and might have reacted differently from the other varieties. These are probably seasonal peculiarities.

Such an experiment as this is, of course, not evaluated by a single season's yields. The statistical methods for perennial crops are described in the next section.

11.16—Perennial plants in randomized blocks. The evaluation of yields from perennial plants is characterized by the continuity of the plot throughout the lifetime of the experiment. Formally, the analysis of variance duplicates that of the plankton data, each record being classified by plot (treatment or variety), by block and by season. The distinction is that the annual readings for a plot are likely correlated from year to year, introducing modifications in the testing of significance.

For illustration, uniformity data on orange crops are given in table 11.37 (15). The investigators reported the yield of each tree each season. On the field plan I delineated blocks of five 16-tree plots, then assigned at random the dummy treatments, $A - E$.

Inspection of the table will reveal several instances of the consistent superiority of one plot over another; for example, the plot designated as variety A in block 2 gave a higher yield during each of the 4 years than that assigned to variety B in the same block. Such permanent differences in fertility between plots contribute to the errors of a variety total and require a corresponding component in the mean square appropriate for testing such totals. The suitable error is the variety-block discrepancy obtained from analysis of the 2-way table shown, each entry being the four-year total. It is as though this were a 5×3 randomized blocks experiment with yields accumulated for its duration. The F -test properly indicates no population differences among the dummy varieties.

The full analysis of the orange data is given in table 11.38 because it throws some light on the test just made and provides others that may be of interest. As before, each two-factor interaction may be tested against season-variety-block, the assumption being that the latter contains only random variation. The fact that the variety-block interaction is highly significant ($F = 222/54 = 4.11$, $F_{.01} = 3.36$) emphasizes what was said above—the fertility differences are scarcely accidents of sampling, and could not safely be ignored in assessing variety yields.

Nonsignificance of the season-variety interaction in an experiment with real varieties would not only be informative but would widen the basis of the inference made about variety differences. To see this, recall the components of variance in the F -ratio for testing variety:

$$\frac{\sigma^2 + 4\sigma_{VB}^2 + 3\sigma_{VS}^2 + 12\sigma_V^2}{\sigma^2 + 4\sigma_{VB}^2}$$

The test is ordinarily made on the assumption that the season effects are population parameters so that inferences apply only to this particular set of seasons (or seasons exactly like them). But if σ_{VS} can be taken as zero, as in this experiment, and if the seasons 1924–27 can be assumed repre-

TABLE 11.37
YIELD (POUNDS PER TREE) OF WASHINGTON NAVEL ORANGES FROM 16-TREE-PLOTS
IN RIVERSIDE, CALIFORNIA

Season	Block	Variety				
		A	B	C	D	E
1924	1	130	152	118	134	140
	2	171	163	164	165	146
	3	159	173	170	164	164
1925	1	132	125	119	119	126
	2	146	138	149	146	126
	3	150	143	140	142	140
1926	1	93	109	85	95	80
	2	122	108	122	123	104
	3	112	120	124	104	121
1927	1	172	158	170	164	160
	2	186	157	166	176	148
	3	158	174	162	161	175

Totals for 4 Seasons

Block	Variety					Mean
	A	B	C	D	E	
1	527	544	492	512	506	129.0
2	625	566	601	610	524	146.3
3	579	610	596	571	600	147.8
Mean	144.2	143.3	140.7	141.1	135.8	141.05

Analysis of Variance for Testing Varieties

Source of Variation	Degrees of Freedom	Mean Square
Variety	4	128
Block	2	
Variety-block	8	222

$$F = 128/222 = 0.58$$

sentative of seasons in general, then the test of $\sigma_v = 0$ may be applied to all such seasons.

The contents of sections 11.13–11.15 should make it plain that testing and interpreting complex experiments cannot be reduced to any simple set of rules. Illustrations have been drawn from the more ordinary types of experiments. If you find it necessary to use less common designs, it may be desirable to consult a mathematical statistician with experience in your

TABLE 11.38
ANALYSIS OF VARIANCE OF ORANGE YIELDS

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Variety	4	513	128
Block	2	4,342	2,171
Variety-block	8	1,780	222
Season	3	28,407	9,469
Season-variety	12	425	35
Season-block	6	1,456	243
Season-variety-block	24	1,306	54
Total	59	38,229	

line of research. N.B.: do this before your experiment starts, not after it is finished.

11.17—Correlation between mean and variance. A fundamental condition for the validity of the F -test of significance is that the two mean squares be independent. If one is conditioned by the other in any manner, the test is invalid. The popularity of analysis of variance has resulted in its application to data in which this condition is not met; specifically to percentages or frequencies of occurrence of some such event as infestation by a parasite. These tend to be distributed in the manner to be described in chapter 16, especially if the event is infrequent. The variance is then closely related to the mean. Hence, if samples are drawn from several such distributions, the mean squares in F will not be independent. This decreases the reliability of the test of significance.

You may be surprised by the suggestion that enumeration data can be treated with methods appropriate to measurement data. Heretofore the two kinds have been kept quite distinct. (For exceptions, recall examples 8.1, 8.4, and 8.8.) There are many instances, however, in which it is either convenient or necessary to pass from one set of methods to the other, as when chi-square was used to test the homogeneity of variance in section 10.13. As experimentation grows more complicated the flexibility of analysis of variance is of increasing advantage, making desirable the transformations that allow its application to enumeration data. Consider for example, a randomized block or latin square experiment containing several plots treated with some disinfectant. Chi-square may be used to test the homogeneity of the percentages of infestation; and the means for two disinfectants may be similarly tested if each is based on plots with uniform infestation. But such results are rare. Even though the plots are treated alike they usually differ significantly in number or percentage of plants attacked. It would not be correct, then, to test the means by use of chi-square. The present trend is toward analyzing the variance of

such enumeration data after a transformation that makes variance independent of means.

Much experimental data, however, whether expressed in percentages or frequencies, may be safely subjected to variance analysis. Percentages which express frequency of occurrence per hundred units when calculated from counts of 100 or more in the *numerator* and when ranging between 20% and 80%, may be expected to yield valid tests in analysis of variance. Similarly, counts of infested plants in experimental plots, for example, running into three figures, usually offer no difficulties.

On the other hand, if percentages result from less than 100 affected individuals, or if the event enumerated is infrequent, some transformation of the variable may be necessary before analysis of variance is carried through. Methods will be discussed in section 16.7.

REFERENCES

1. F. E. ALLAN and J. WISHART. *Journal of Agricultural Science*, 20:399 (1930).
2. E. R. BECKER and PHOEBE R. HALL. *Parasitology*, 25:397 (1933).
3. A. E. BRANDT. Thesis submitted for the degree Doctor of Philosophy, Iowa State College (1932).
4. BERNICE BROWN. *Proceedings of the Iowa Academy of Science*, 38:205 (1932).
5. WILLIAM G. COCHRAN and GERTRUDE M. COX. *Experimental Designs*. Mimeographed copy (1944).
6. DWIGHT ESPE. *Data of the Iowa Agricultural Experiment Station*.
7. R. A. FISHER. *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh (1925-1944).
8. R. A. FISHER. *The Design of Experiments*. Oliver and Boyd, Edinburgh (1935-1942).
9. H. C. FORSTER and A. J. VASEY. *Journal of the Department of Agriculture of Victoria, Australia*, 30:35 (1932).
10. JOHN W. GOWEN. *Data as yet unpublished*. Iowa Agricultural Experiment Station (1945).
11. KATHERINE L. HOFFMAN. Thesis submitted for the degree of Master of Science, Iowa State College (1939).
12. F. R. IMMER, H. K. HAYES, and LEROY POWERS. *Journal of the American Society of Agronomy*, 26:403 (1934).
13. H. W. LI, C. J. MENG, and T. N. LIU. *Journal of the American Society of Agronomy*, 28:1 (1936).
14. BELLE LOWE. *Proceedings of the seventh world's poultry congress and exposition*, page 500 (1939).
15. E. R. PARKER and L. D. BATCHELOR. *Hilgardia*, 7:81 (1932).
16. R. H. PORTER. *Cooperative Soybean Seed Treatment Trials*, Iowa State College Seed Laboratory (1943).
17. F. B. SMITH. *Data from the Iowa Agricultural Experiment Station* (1936).

18. GEORGE W. SNEDECOR. *Journal of the American Statistical Association*, 29:389 (1934).
19. GEORGE W. SNEDECOR and GERTRUDE M. COX. *Iowa Agricultural Experiment Station Research Bulletin* 180 (1935).
20. GEORGE W. SNEDECOR and C. C. CULBERTSON. *Record of Proceedings, American Society of Animal Production* (1932).
21. NORMAN V. STRAND and RAYMOND J. JESSEN. *Iowa Agricultural Experiment Station Research Bulletin* 315 (1943).
22. "STUDENT." *Biometrika*, 29:363 (1938).
23. O. TEDIN. *Journal of Agricultural Science*, 21:191 (1931).
24. C. P. WILSIE. *Data from Iowa State College Agricultural Experiment Station* (1944).
25. C. P. WINSOR and G. L. CLARKE. *Sears Foundation: Journal of Marine Research*, 3:1 (1940).
26. F. YATES. *The Empire Journal of Experimental Agriculture*, 1:129 (1933).
27. F. YATES. *Journal of Agricultural Science*, 23:108 (1933).
28. F. YATES. *Supplement to the Journal of the Royal Statistical Society*, 2:181 (1935).
29. F. YATES. *Journal of Agricultural Science*, 26:301 (1936).
30. F. YATES and W. G. COCHRAN. *Journal of Agricultural Science*, 28:556 (1938).

Chapter 12

TWO VARIATES IN TWO OR MORE GROUPS. COVARIANCE

12.1—Introduction. Covariance is a term applied to a statistical method used to reduce the data from experiments involving two or more variates measured in several groups. In this chapter the method will be applied to two variates only. In that respect we shall be occupied in extending the methods of regression and correlation as developed in chapters 6 and 7 to include more than one group of data. In the use of correlated data we shall be broadening the methods of analysis of variance treated in the last two chapters. The use of the word covariance to designate a statistical method as well as a mean product (section 6.5) need cause no confusion.

12.2—An experiment with a single criterion of classification. The rats whose food intake and gain in weight are recorded in table 12.1 were

TABLE 12.1

FOOD INTAKE, X (10-CALORIE UNITS), AND GAIN IN WEIGHT, Y (GRAMS), OF 60 RATS
RECEIVING SIX RATIONS IN LOTS OF 10

Rat	Lot											
	1		2		3		4		5		6	
	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y
1	108	73	99	98	194	94	165	90	124	107	140	49
2	136	102	117	74	198	79	164	76	95	95	177	82
3	138	118	90	56	196	96	161	90	116	97	189	73
4	159	104	141	111	198	98	159	64	112	80	142	86
5	146	81	106	95	210	102	175	86	123	98	216	81
6	141	107	112	88	196	102	135	51	110	74	200	97
7	175	100	110	82	230	108	132	72	137	74	255	106
8	149	87	117	77	222	91	190	90	105	67	173	70
9	174	117	111	86	220	120	145	95	135	89	153	61
10	176	111	122	92	228	105	142	78	126	58	160	82
Sums	1,502	1,000	1,125	859	2,092	995	1,568	792	1,183	839	1,805	787
SX^2	229,760		128,245		439,544		248,886		141,525		337,433	
SY^2	102,062		75,819		100,075		64,462		72,613		64,401	
SXY	151,846		97,776		208,892		125,370		99,195		145,872	

Experiment totals: $SX = 9,275$ $SY = 5,272$ $SXY = 828,951$
 $SX^2 = 1,525,393$ $SY^2 = 479,432$

classified according to the single criterion, ration fed. The object is to learn if the six foods were associated with differences in gain. It was recognized that such differences might be caused either by unequal availability of the nutritive elements in the food or by variations in palatability. The investigation of the relations between two variates such as food intake and gain, occurring in two or more groups, calls for the methods of covariance.

We shall first ask these questions: did the six rations induce significant differences among the lot gains, and how were these gains conditioned by food intake? The first question is easily answered by the straightforward analysis of variance in the manner of chapter 10. Table 12.2 indicates

TABLE 12.2
ANALYSIS OF VARIANCE OF RAT GAINS IN WEIGHT

Source of Variation	Degrees of Freedom	Mean Square
Rats within lots	54	215
Rations	5	923**

$$F = 923/215 = 4.3$$

highly significant differences among the lot gains. The answer to the next question is the subject matter of this chapter on covariance.

12.3—The test of significance of differences among adjusted group means. Since this test is of general utility and fairly easy to make, the mechanism will be presented now leaving explanations till later. The first step consists in the calculation of the sum of squares of errors of estimate for the experiment as a whole, using the method of chapter 6. The experiment totals at the bottom of table 12.1 lead to the following:

Correction terms: For X , $(9,275)^2/60 = 1,433,760.42$
 For Y , $(5,272)^2/60 = 463,233.07$
 For XY , $(9,275)(5,272)/60 = 814,963.33$

Sums of squares: $Sx^2 = 1,525,393 - 1,433,760.42 = 91,632.58$
 $Sy^2 = 479,432 - 463,233.07 = 16,198.93$

Sum of products: $Sxy = 828,951 - 814,963.33 = 13,987.67$

The sums of squares and products, entered in the first line of table 12.3, are used to compute the total sum of squares of errors of estimate (section 6.8):

$$Sy^2 - (Sxy)^2/Sx^2 = 16,198.93 - (13,987.67)^2/91,632.58 = 14,063.72$$

The degrees of freedom are two less than the entire number of rats.

TABLE 12.3
ANALYSIS OF COVARIANCE AND TEST OF SIGNIFICANCE OF ADJUSTED LOT MEANS

Source of Variation	Degrees of Freedom	Sums of Squares and Products			Errors of Estimate		
		Sx^2	Sxy	Sy^2	Sum of Squares ¹	Degrees of Freedom	Mean Square
Total Rations	59	91,632.58	13,987.67	16,198.93	14,063.72	58	
	5	67,662.68	5,520.97	4,612.93			
Within lots (error)	54	23,969.90	8,466.70	11,586.00	8,595.37	53	162.2
For test of significance of adjusted means					5,468.35	5	1,093.7**
$^1 Sy^2 - \frac{(Sxy)^2}{Sx^2} \qquad F = 1,093.7/162.2 = 6.74$							

We next compute sums of squares and products characterizing the regression of the lot mean gain on lot mean food intake. The calculations are in the forms typical of analysis of variance, employing the lot sums:

$$\text{Food intake: } \frac{(1,502)^2 + \dots + (1,805)^2}{10} - 1,433,760.42 = 67,662.68$$

$$\text{Gains: } \frac{(1,000)^2 + \dots + (787)^2}{10} - 463,233.07 = 4,612.93$$

$$\text{Products: } \frac{(1,502)(1,000) + \dots + (1,805)(787)}{10} - 814,963.33 = 5,520.97$$

These quantities entered in line 2 might be used to calculate the sum of squares of errors of estimate of lot mean gains measured from the regression of means. That will be done later, but it is unnecessary for the present test. Instead, the sums of squares and products are merely subtracted from the corresponding quantities in line 1 to get those of line 3. These last can be calculated directly from the data of the individual lots, in the manner of table 12.6, thus completely verifying the computations. At present we use them in a regression *within lots* similar to the *total* first calculated. As will be shown later, this regression is an average of the individual lot regressions. From it have been eliminated any effects of the differences among the lot means themselves. The associated sum of squares of errors of estimate, 8,595.37, is calculated as usual. Since this sum is made up of uncontrolled variations in gain *after due allowance is made for food intake*, it is the basis for an estimate of experimental error. The accompanying degrees of freedom are $54 - 1 = 53$, the one deducted belonging to the last regression computed, that within lots.

The difference in the third-to-last column of the table, divided by the corresponding degrees of freedom, $5,468.35/5 = 1,093.7$, is the mean square appropriate for testing the significance of the differences among the adjusted lot mean gains. Since F is highly significant the conclusion is that the quantity of food intake does not explain the differences among the lot gains: after these gains are adjusted to a common food intake basis they still differ significantly. Evidently the nutritive materials in these foods are not equally adapted to assimilation by the rats. Without the covariance analysis, one would not know whether the differences in gain were attributable to quantity of nutritive materials eaten or to their quality.

One might or might not get similar results by first reducing each gain to a per-calorie basis, then analyzing the variance of *gain per calorie*. Unless previous investigations have established the validity of the analysis of the ratios, it is safer to use the covariance method. Besides, it furnishes a wealth of information which would be lost if only gain per calorie were considered. This will be increasingly evident as the subject unfolds.

EXAMPLE 12.1—This is a set of numbers designed for easy computation:

Item	Group							
	1		2		3		4	
	X	Y	X	Y	X	Y	X	Y
1	29	22	15	30	16	12	5	23
2	20	22	9	32	31	8	25	25
3	14	20	1	26	26	13	16	28
4	21	24	6	25	35	25	10	26
5	6	12	19	37	12	7	24	23

Verify the following:

Source of Variation	Errors of Estimate	
	Degrees of Freedom	Mean Square
Error	15	16.7
For testing significance of adjusted means	3	307.7

12.4—The adjusted lot means and a question in design. It is instructive to calculate the adjusted lot mean gains, then compare them with the unadjusted. You may recall the formula for the adjusted values (section 6.7), $Y - bx$. Compute the error regression coefficient from table 12.3,

$$b = \frac{S_{xy}}{S_x^2} = \frac{8,466.70}{23,969.90} = 0.3532 \text{ grams per unit}$$

From table 12.1, $\bar{x} = 9,275/60 = 154.6$ units, while the lot sums readily yield the lot means. The resulting information is put together in table 12.4. In the main the adjusted and actual mean gains are not greatly

TABLE 12.4
CALCULATION OF ADJUSTED MEAN GAINS OF RATS
 $\bar{x} = 154.6$ grams $b = 0.3532$ grams per unit food

Lot	Mean Food Intake $\bar{X}$	Deviation From Experiment Mean $\bar{x}$	Product $b\bar{x}$	Mean Gain $\bar{Y}$	Adjusted Mean Gain $\bar{Y} - b\bar{x}$
1	150.2	- 4.4	- 1.6	100.0(1) *	101.6(1) *
2	112.5	-42.0	-14.8	85.9(3)	100.7(2)
3	209.2	54.6	19.3	99.5(2)	80.2(4)
4	156.8	2.2	0.8	79.2(5)	78.4(5)
5	118.3	-36.3	-12.8	83.9(4)	96.7(3)
6	180.5	25.9	9.1	78.7(6)	69.6(6)
Sums	927.5	0.0	0.0	527.2	527.2

* Rank.

discrepant. The high- and low-ranking lots are the same in both lists. The major shift is in the lots ranking third. With a knowledge of the food compositions, the nutritionist might draw some conclusions from the changes in rank; the statistician can only point them out.

The graphical representation of figure 12.1 may be of assistance.

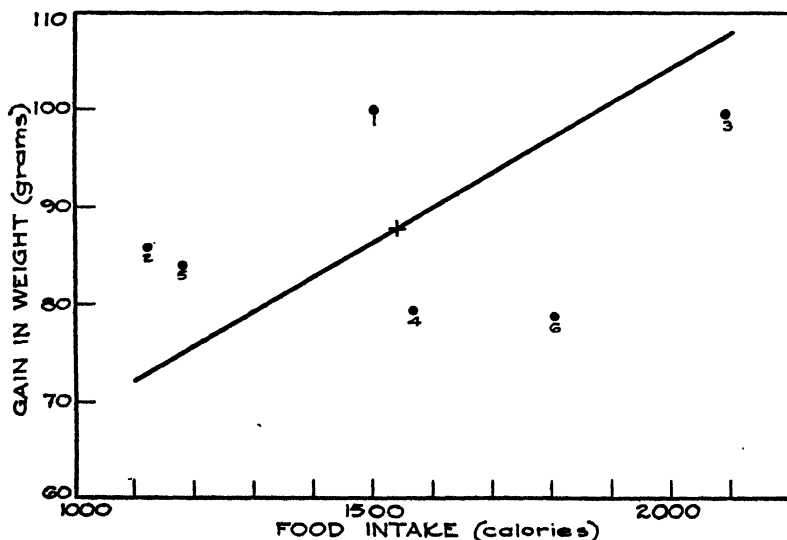


FIG. 12.1—Average regression of gain on food intake in six lots of rats. The dots represent the lot means.

The regression equation *within lots* is

$$\begin{aligned}\hat{Y} &= 87.87 + 0.3532 (X - 154.6) \\ &= 0.3532X + 33.27\end{aligned}$$

This is plotted in the figure along with the six points representing lot means. What we have learned is that these points deviate significantly from the line. The amounts of food consumed do not account for the lot gains.

The investigator will wish to know if it has been worthwhile to introduce the independent variate into the experiment. Three items of information will assist him in making judgment. First, the list of actual and adjusted means. Sometimes the rank order of adjusted means is quite different from that of the unadjusted, and the shifts may be interpreted. Second, a comparison of the sum of squares of errors of estimate (table 12.3) used for testing significance, 5,468.35, with $Sy^2 = 4,612.93$. In this example the former is greater than the latter, but the opposite may occur. Any notable contrast in the sizes of these sums of squares is informative. Third, the change in the precision of the experiment owing to adjustment of the error sum of squares. The changes taking place in this experiment are clearly indicated in table 12.5. The sum of squares, $Sy^2 = 11,586$,

TABLE 12.5
ANALYSIS OF ERROR VARIANCE IN RAT DATA

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Within lots, unadjusted gains, Sy^2	54	11,586	215
Reduction due to regression, $(Sxy)^2/Sx^2$	1	2,991	2,991 **
Error for adjusted gains	53	8,595	162

$$F = 2,991/162 = 18$$

with 54 degrees of freedom is analyzed into two parts, one with a single degree of freedom measuring the variation attributable to linear regression, the other with 53 *d.f.* being assigned to error. Not only is the mean square for error reduced from 215 to 162, but the reduction in sum of squares due to regression is highly significant. This reveals a valuable increase in the precision of the experiment. A consideration of these three features of the covariance method will enable the investigator to decide whether to retain the independent variable in similar experiments.

EXAMPLE 12.2—In the following table are initial weights X (pounds) and average daily gains Y (pounds per day) of four lots of swine fed different rations. For ease in computation these data are somewhat oversimplified. (You will remember that there should be ranges of from 25 to 40 in the significant digits of each variate.) Make the necessary calculations to test the significance of the differences among the lot mean

gains ($F = 15.5$) as well as among the adjusted mean gains ($F = 19$). For your guidance, some results are appended to the data.

Pig Number	Lot 1		Lot 2		Lot 3		Lot 4	
	Weight $\bar{X}$	Gain $\bar{Y}$	Weight $\bar{X}$	Gain $\bar{Y}$	Weight $\bar{X}$	Gain $\bar{Y}$	Weight $\bar{X}$	Gain $\bar{Y}$
1	62	1.2	79	2.0	71	1.2	61	1.4
2	73	1.4	65	1.8	60	1.3	59	1.8
3	58	1.3	57	1.6	54	1.4	59	1.6
4	43	1.3	51	1.8	50	1.4	53	1.5
5	50	1.4	57	1.9	60	1.2	56	1.7
6	44	1.2	66	1.5	61	1.2	50	1.5
7	48	1.0	44	1.6	44	1.2	45	1.4
8	51	1.6	41	1.5	53	1.0	39	1.4
9	40	1.2	44	1.8	41	1.1	38	1.3
10	38	1.1	36	1.3	38	1.1	45	1.3
	507	12.7	540	16.8	532	12.1	505	14.9

Source of Variation	Degrees of Freedom	Sums of Squares and Products			Errors of Estimate		
		Sx^2	Sxy	Sy^2	Sum of Squares	Degrees of Freedom	Mean Square
Total Lots	39	4,355.60	37.55	2.464	2.1403	38	
	3	93.40	3.61	1.389			
Error	36	4,262.20	33.94	1.075	0.8047	35	0.023
Difference for testing adjusted means					1.3356	3	0.445**

$$F = 0.445/0.023 = 19$$

EXAMPLE 12.3—The data in example 12.2 are taken from an experiment (3) in which the attempt was made to forecast the gains of the pigs as a basis for assignment to the lots. The four pigs numbered 1 were expected to gain at the same rate if they had been fed the same ration. That is, the pig numbers roughly represented levels of gaining ability. One is warranted, then, in introducing a second criterion of classification, making the experiment essentially a randomized block design like that of section 11.2. The sum of the four gains in each of the 10 lines constitutes the block sum. It is easy to compute this table:

Source of Variation	Degrees of Freedom	Mean Square
Blocks or outcome groups	9	0.0558*
Lots or rations	3	0.4630**
Error	27	0.0212

$$F = 0.4630/0.0212 = 22$$

You can now make some comparisons between the two designs for experiments. If you know enough about your experimental material to classify it into outcome groups, you can segregate the sum of squares for such groups by the methods of chapter 11,

thus avoiding covariance. You will notice that the mean square for outcome groups is significant; also that F for lots is larger than in the table of example 12.2. On the other hand, the error mean square now has only 27 $df.$, and you have sacrificed all the extra information available from the covariance analysis. Both designs are excellent. In situations like the one under discussion, the investigator may choose between them. The object of this example is to contrast the possibilities which they offer.

12.5—Covariance in the groups. In experiments expressly designed for covariance, the foregoing test of significance is usually the end of the statistical investigation. For two reasons, however, it is worthwhile to examine somewhat minutely the covariance in the several groups, together with the relations among these groups. First, it is desirable to gain a knowledge of the structure of the test we have just learned to make. In the second place, the statistician is often called upon to analyze sets of data in which covariance had not been contemplated. The relations among the groups of data may be more pertinent than any single test of significance.

Let us return to the rat data in table 12.1 and calculate the regression in each lot. The results are entered in table 12.6. In each line the devia-

TABLE 12.6
REGRESSION AND CORRELATION DATA IN SIX LOTS OF RATS

Lot	Degrees of Freedom	Sums of Squares and Products			Correlation Coefficient ¹	Regression Coefficient ²	Errors of Estimate	
		Sx^2	Sxy	Sy^2			Sum of Squares ³	$df.$
1	9	4,159.60	1,646.00	2,062.00	0.5620	0.3957	1,410.66	8
2	9	1,682.50	1,138.50	2,030.90	0.6159	0.6767	1,260.51	8
3	9	1,897.60	738.00	1,072.50	0.5173	0.3889	785.48	8
4	9	3,023.60	1,184.40	1,735.60	0.5170	0.3917	1,271.65	8
5	9	1,576.10	-58.70	2,220.90	-0.0314	-0.0372	2,218.71	8
6	9	11,630.50	3,818.50	2,464.10	0.7133	0.3283	1,210.42	8
Sums 54		23,969.90	8,466.70	11,586.00	0.5081	0.3532	8,157.43	48

$$^1 Sxy / \sqrt{(Sx^2)(Sy^2)}$$

$$^2 Sxy / Sx^2$$

$$^3 Sy^2 - (Sxy)^2 / Sx^2$$

tions whose sums of squares and products are set down are measured from the lot mean, not from the experiment mean as in table 12.3. That implies a set of corrections for each lot, the set for lot 1 being:

$$\text{For } X: (1,502)^2/10 = 225,600.40$$

$$\text{For } Y: (1,000)^2/10 = 100,000.00$$

$$\text{For } XY: (1,502)(1,000)/10 = 150,200.00$$

The sums of squares and products for lot 1 follow from applying these corrections to the totals in the last three lines of table 12.1:

$$\text{For lot 1: } Sx^2 = 229,760 - 225,600.40 = 4,159.60$$

$$Sy^2 = 102,062 - 100,000.00 = 2,062.00$$

$$Sxy = 151,846 - 150,200.00 = 1,646.00$$

These and similar results are entered in the table.

Since we are going into the data rather thoroughly, it is of interest to compute correlation and regression coefficients as well as the sums of squares of errors of estimate. The values of each set of coefficients differ considerably among themselves. One of our jobs is to learn if these differences are significant, or if they represent merely sampling variation. Another is to consider a way of averaging these coefficients. At the moment, let us compute and plot in figure 12.2 the regression equations for the six lots. That for lot 1 is

$$\begin{aligned}\hat{Y} &= 100.0 + 0.3957(X - 150.2) \\ &= 0.3957X + 40.57,\end{aligned}$$

the means coming directly from the sums in table 12.1. We are to learn if the slopes of these lines differ significantly.

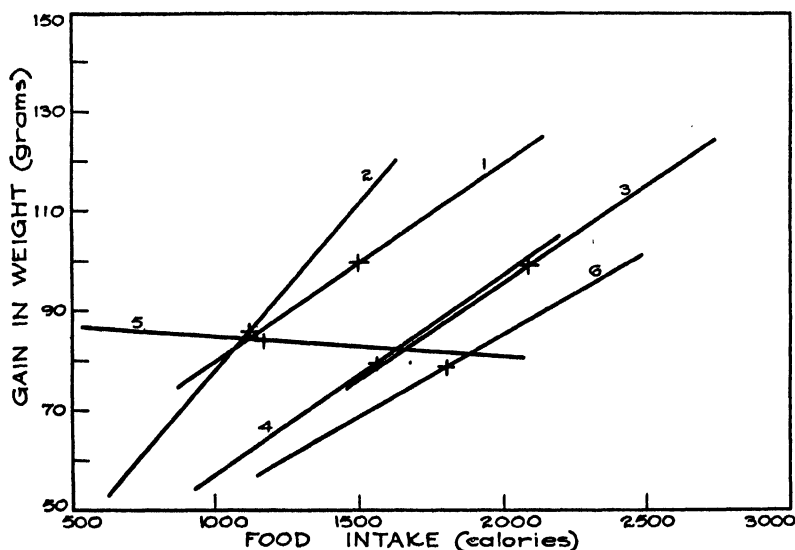


FIG. 12.2—Regressions of gain on food intake in six lots of rats. The average of these regressions is plotted in figure 12.1.

A notable feature of table 12.6 is this: the totals in the three columns Sx^2 , Sxy , and Sy^2 are the same as the differences in the third line of table 12.3. This fact furnishes a perfect verification of the calculations in the two tables. But, more important, it gives a clue to the character of the sums of squares and products within lots in the earlier table. It is now evident that they are the pooled values from all the groups, each deviation involved being measured from its own group mean. Hence, the *within lot* regression of table 12.3 is an average of the six lot regressions of table 12.6. For comparison, note that this regression, computed in section 12.4, is 0.3532 grams per unit food intake, intermediate in value among the lot

regressions. Thus we have been successful in our quest for an average of the individual regressions of table 12.6.

An interesting contrast appears from a comparison of the *within lot* sum of squares of errors of estimate in table 12.3 with the sum of those in the last-but-one column of table 12.6. They are put in juxtaposition in table 12.7. Those in table 12.6 are sums of squares of deviations from

TABLE 12.7
ANALYSIS OF ERRORS OF ESTIMATE FROM AVERAGE REGRESSION WITHIN GROUPS

Source of Variation	Degrees of Freedom	Errors of Estimate	
		Sum of Squares	Mean Square
Deviations from average (error) regression within lots, table 12.3	53	8,595.37	
Deviations from individual lot regressions, table 12.6	48	8,157.43	169.9
Differences among lot regressions	5	437.94	87.6

individual regressions in the several lots. According to the principle of least squares each of these sums is less than that from any other straight line; hence, each is less than the sum of squares of the deviations from the average regression of table 12.3. That explains why the second sum in table 12.7 is less than the first. It turns out that the difference between them, 437.94, with 5 degrees of freedom is associated with the variation of the lot regression coefficients. The corresponding mean square, 87.6, is to be tested for significance. For the test, the mean square deviation from individual lot regressions is a proper comparison. This mean square is an average of that variation in gains not explained either by ration differences or by the individual regressions, analogous to discrepancy in a randomized block experiment. It is therefore attributable to experimental error. When you attempt the test, you find that the mean square for differences among regressions is actually less than error. The test is quite unnecessary, the lot regressions being more nearly alike than would be expected half the time in sampling from a homogeneous population.

We have thus reached the second of our immediate objectives, a test of the significance of the differences among the group regression coefficients. In a sense, we got the trailer in front of the car, averaging the regression coefficients before we showed that they do in fact belong to a common population. Since the average turns out to be legitimate, we may revise our opinions to some extent. We may properly look upon the average regression as the best available estimate of the population regression and may leave behind the individual regressions as being merely items in the larger sample. For further testing we may, as in table 12.3,

use the mean square 162.2 as the estimate of error. Active interest in the individual regressions almost vanishes with proof that they do not differ significantly.

Having arrived at such conclusions, we may consider quite properly an average of the correlation coefficients in table 12.6. Since they may be samples from a homogeneous population, their average is computed from the sums of squares and products associated with the average regression, the result being 0.5081, the same as may be got by section 7.6. This is larger than the experiment (total) correlation, 0.3631, chiefly because of eliminating from the denominator the large sum of squares attributable to variation of the lot means.

This detailed examination into the individual lot regressions yields two more bits of information sometimes well worth while. We return to table 12.3 and compute the regression averages for the ration means:

1. $r = 5,520.97 / \sqrt{(67,662.68)(4,612.93)} = 0.3125$
2. $b = 5,520.97 / 67,662.68 = 0.08160$ grams per unit food
3. Sum of squares of errors of estimate
 $= 4,612.93 - (5,520.97)^2 / 67,662.68 = 4,162.44$

The small reduction in sum of squares shows that there is no pronounced trend. Lot mean gains do not increase regularly with mean food intake, but are decidedly erratic. The regression equation for lot means is

$$\hat{Y} = 87.87 + 0.08160 (X - 154.6) \\ = 0.08160X + 75.25$$

This equation may be plotted in figure 12.1 for contrast with the average regression and the lot means. It is easy to see, however, that the lot means follow this regression no better than that within lots. A formal test is supplied in table 12.8. Here is entered a sum of squares of errors of estimate computed in each of the first three lines of table 12.3. The mean square for lot means is highly significant as compared with the average regression within lots. This verifies the conclusion reached above from an examination of the reduction in sum of squares. The variation in these lot mean gains is independent of their regression.

TABLE 12.8
ANALYSIS OF ERRORS OF ESTIMATE FROM THREE REGRESSIONS

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Total from table 12.3	58	14,063.7	
Lot means	4	4,162.4	1,040.6**
Average within lots, table 12.3	53	8,595.4	162.2
Remainder	1	1,305.9	1,305.9**

There remains a single degree of freedom in table 12.8, with an associated mean square, 1,305.9, that occasionally has an interesting interpretation. It represents the difference between two regressions, that of lot means (0.0816) and the average within lots (0.3532). If the mean square in the second line of the table had been nonsignificant, this difference might have had a clear meaning. Its significance would have indicated that the gains of individuals within the groups followed a trend different from that of group means. In the present experiment where there is no pronounced trend of means, it is idle to bother about the difference between this and any other trend. Only if there are well defined trends both among individuals within lots and among lot means is this degree of freedom likely to yield interpretable information.

EXAMPLE 12.4—The analysis of the errors of estimate for example 12.1 is:

Source of Variation	Degrees of Freedom	Mean Square
Deviations from individual regressions	12	16.32
Regression coefficients	3	18.40
Group means from their own regression	2	10.60
Remainder	1	901.86

For an explanation, represent graphically the original data.

EXAMPLE 12.5—Complete the covariance analysis of the lots of pigs in example 12.2. The results are:

Source of Variation	Degrees of Freedom	Mean Square Errors of Estimate
Within lots, pooled regression	35	0.0230
Within lots, individual lot regressions	32	0.0235
Lot regression coefficients	3	0.0179
Lot means	2	0.6248 **
Difference between regressions within lots and between lot means	1	0.0861

This indicates uniformity of behavior in all parts of the experiment save only in lot means. Their differences are evidently the result of the rations, being unexplained by variations in initial weight. This is merely additional information—the proper test is that of example 12.2.

EXAMPLE 12.6—For the pigs compute the adjusted lot means, 1.28, 1.67, 1.20, 1.50.

EXAMPLE 12.7—Calculate the regression within lots of pigs,

$$\hat{Y} = 0.00796X + 0.997,$$

and plot it on the same graph with the lot means, showing the manner in which these means depart from regression.

TABLE 12.11
TEST OF SIGNIFICANCE OF DIFFERENCE BETWEEN ADJUSTED MEAN GAINS
AT TWO LEVELS OF PROTEIN

Source of Variation	Errors of Estimate		
	Degrees of Freedom	Sum of Squares	Mean Square
Total from table 12.3	58	14,063.72	
Average within levels	57	11,233.08	197
Between adjusted means	1	2,830.64	2,831**

Adjusted mean gains: high level, 94.7; low level, 81.0

two levels of protein intake differ significantly, but the regressions are remarkably similar.

A striking contrast between tables 12.12 and 12.8 is the absence of the two sources of variation, *lot means* and *remainder*. The *remainder* is, in fact, the single degree of freedom *between regression coefficients* in table 12.12. There can be no deviation from the regression of means—since there are only two of them they determine the two points that fix the regression. There is neither degree of freedom nor sum of squares for testing this regression.

EXAMPLE 12.8—Compute the regression equations for the two levels:

$$\text{high: } \hat{Y} = 0.1734X + 67.82$$

$$\text{low: } \hat{Y} = 0.0897X + 66.97$$

Plot these on a graph, then plot the six lot means on the same graph.

12.7—Covariance in a randomized block experiment. When it is necessary to use statistical control in a randomized block experiment, the test of significance of the differences among the adjusted means is easily made. In table 12.13 are data from an experiment of this type. The

TABLE 12.12
TEST OF SIGNIFICANCE BETWEEN REGRESSION COEFFICIENTS AT HIGH AND
LOW LEVELS OF PROTEIN

Source of Variation	Errors of Estimate		
	Degrees of Freedom	Sum of Squares	Mean Square
Average within levels	57	11,233	
Deviations from individual level regressions	56	11,081	198
Between regression coefficients	1	152	152

TABLE 12.13
NUMBERS OF BEETS PER PLOT AND YIELDS (TONS PER ACRE) IN A
RANDOMIZED BLOCK EXPERIMENT

Fertilizer Applied	Number and Yields	Block						Treat- ment Sums
		1	2	3	4	5	6	
None	Number	183	176	291	254	225	249	1,378
	Yield	2.45	2.25	4.38	4.35	3.42	3.27	20.12
Superphos- phate, <i>P</i>	Number	356	300	301	271	288	258	1,774
	Yield	6.71	5.44	4.92	5.23	6.74	4.74	33.78
Muriate of potash, <i>K</i>	Number	224	258	244	217	192	236	1,371
	Yield	3.22	4.14	2.32	4.42	3.28	4.00	21.38
<i>P</i> + <i>K</i>	Number	329	283	308	326	318	318	1,882
	Yield	6.34	5.44	5.22	8.00	6.96	6.96	38.92
<i>P</i> + sodium nitrate, <i>N</i>	Number	371	354	352	331	290	410	2,108
	Yield	6.48	7.11	5.88	7.54	6.61	8.86	42.48
<i>K</i> + <i>N</i>	Number	230	221	237	193	247	250	1,378
	Yield	3.70	3.24	2.82	2.15	5.19	4.13	21.23
<i>P</i> + <i>K</i> + <i>N</i>	Number	322	367	400	333	314	385	2,121
	Yield	6.10	7.68	7.37	7.83	7.75	7.39	44.12
Block sums	Number	2,015	1,959	2,133	1,925	1,874	2,106	12,012
	Yield	35.00	35.30	32.91	39.52	39.95	39.35	222.03

object of the investigation was to learn the effect of certain fertilizers on the yield of sugar beets. But the number of plants differed from plot to plot. It was necessary, therefore, to examine the effect of the variable stand and to estimate the yields on the basis of equal numbers of beets. For the two variates, number of beets X and yield Y , the sums of the recorded values are entered in the table. For a complete check, these should be run up for both treatments and blocks. From the experiment sums the three corrections for mean are calculated:

For X : 3,435,432

For Y : 1,173.7457

For XY : 63,500.58

From the experiment sums of squares and products, the values of Sx^2 , Sxy , and Sy^2 are computed and entered in the first line of table 12.14. The corresponding results for blocks in line 2 are:

$$Sx^2 = \frac{(2,015)^2 + \dots + (2,106)^2}{7} - 3,435,432 = 7,472.57$$

$$S_y^2 = \frac{(35.00)^2 + \dots + (39.35)^2}{7} - 1,173.7457 = 6.3134$$

$$S_{xy} = \frac{(2,015)(35.00) + \dots + (2,106)(39.35)}{7} - 63,500.58 = -116.56$$

The third line, treatments, comes from similar computations applied to the treatment sums, the denominator now being 6, the number of items entering into each sum. Next the numbers in lines 2 and 3 are deducted from the line of totals. The remainder corresponds to the variation commonly ascribed to error, the discrepancy among the adjusted plot yields.

TABLE 12.14
ANALYSIS OF COVARIANCE OF STAND AND YIELD OF SUGAR BEETS, TOGETHER
WITH TEST OF SIGNIFICANCE

Source of Variation	Degrees of Freedom	Sums of Squares and Products			Errors of Estimate		
		S_x^2	S_{xy}	S_y^2	Sum of Squares	Degrees of Freedom	Mean Square
Total	41	152,158.00	4,163.69	142.4022			
Blocks	5	7,472.57	-116.56	6.3134			
Treatments	6	116,020.33	3,598.05	112.8562			
Error	30	28,665.10	682.20	23.2326	6.9969	29	0.241
Treatment + error	36	144,685.43	4,280.25	136.0888	9.4655	35	
Difference for testing adjusted treatment means					2.4686	6	0.411
$F = 0.411/0.241 = 1.71, \quad P_{.05} = 2.43$							

Let us pause to analyze the variance of the yields as in table 12.15. The block differences appear nonsignificant, apparently indicating unusually uniform field conditions. The yields, however, varied decidedly with treatment. We wish to know how these yields are related to stand.

Returning to table 12.14, the next step is different from any taken before. Since the total in line 1 has been subdivided into more than two

TABLE 12.15
ANALYSIS OF VARIANCE OF YIELDS Y OF SUGAR BEETS WITH SEVEN TREATMENTS
Data taken from table 12.14

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Blocks	5	6.3134	1.26
Treatments	6	112.8562	18.81**
Error	30	23.2326	0.774

parts, it is necessary to form a new total containing only the two effects we are interested in—treatment and error. This new total is in the line, *treatment + error*, degrees of freedom, sums of squares and products all being added to get the appropriate numbers. The remainder of the testing is like that of table 12.3. Sums of squares of errors of estimate are calculated by use of the usual formula in lines *error* and *treatment + error*. The former is subtracted from the latter to yield the sum of squares, 2.4686, appropriate for testing significance. You see that the adjusted treatment means do not differ significantly. When this result is compared with that of table 12.15, it is clear that the yield differences that seemed to be due to treatment were largely the result of differences in stand. The average weight of the individual beets wasn't greatly affected by the fertilizers.

There is the possibility that the differences in stand were themselves the result of the treatments. For evidence, it is worth completing the analysis of the variance of numbers of beets, X , as is done in table 12.16.

TABLE 12.16

ANALYSIS OF VARIANCE OF NUMBERS OF PLANTS PER PLOT IN SUGAR BEET EXPERIMENT
Data taken from table 12.14

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Blocks	5	7,472.57	1,495
Treatments	6	116,020.33	19,337**
Error	30	28,665.10	956

Two pertinent facts appear. The block stands did not differ significantly, indicating that the differential survival of the plants was not a soil characteristic. On the other hand, the stands in the plots differed significantly with treatment. We thus come back to the starting point of our investigation; the yields are affected by treatment, but we now have the valuable information that the result is brought about chiefly by treatment effects on stand rather than on individual beet weight.

Another question of interest, referred to above, is this: did the treatments affect the weight of the individual beets? If you divide each yield in table 12.13 by the corresponding number of plants, then analyze the results, you will get highly significant differences among the treatment means, most of it associated with the contrast between the plots receiving superphosphate and those without. (See, also, example 12.10.) Furthermore, a covariance analysis of mean beet weight *adjusted for number per plot* shows significant differences, again attributable largely to superphosphate. There is little doubt that in this experiment the superior yields on the plots with superphosphate are attributable to both numbers and weights of the plants produced, yield being the product of the two.

Distinction should be drawn sharply between the uses of covariance analysis in the two situations where the treatment mean $\bar{X}$'s do and do not

differ significantly. The latter circumstance characterized the pig initial weights of example 12.3, the object of the covariance analysis being to gain increased precision in the test of significance. Significant differences among the means of the independent variable, found in both the beet numbers and the food intake of section 12.3, focus attention on the causes and conditions of the variations in Y . Did the rats gain differently because of the differences in food intake, or in spite of them? If one assumes equality in stand among plots of beets, would the yields (individual beet weights) still give evidence of being different? Physiological questions then come up, such as the balance between competition and food supply in the plots with varying stands.

In examining the effect of stand on yield in maize, as another example, significant differences in stand may be traced to differential viability in seed (associated perhaps with variety) or to experimental methods of planting. An investigation of the latter kind may not require covariance at all, the knowledge of differences in stand being the object of the experiment; but the physiologist likes to know if decreased stand was compensated by increased yield per plant. The problem of viable seed will interest the breeder and dealer as well as the farmer. Some part of the covariance analysis contains information for each.

The reader will be well repaid by a critical examination of the nature of the data in the line for *error*, table 12.14. Of what errors of estimate, for example, is 6.9969 the sum of squares? The error regression equation of the foregoing paragraph might be used for calculating 42 errors of estimate, the sum of whose squares is actually 6.9969. Yet little light would be shed on the fundamental character of the error regression. This equation really relates yield to stand independently of both block and treatment. Let us look at the mechanism. But first refresh your memory of expected values in section 11.3. Following those methods one could supply expected numbers of beets and also expected yields in each of the 42 cells of table 12.13. These would lead to two complete sets of deviations from expected values. It is these deviations that supply the key to the meaning of the *error* regression statistics. They are independent of block and treatment, being the remainders from those very expected values which are determined by block and treatment constants. If one correlates the 42 pairs of deviations in the manner of chapter 7, he will get the same result as may be got from table 12.14,

$$682.20/\sqrt{(28,665.10)(23.2326)} = 0.8360$$

The regression of the yield deviation on the number-of-beets deviation is the same as $682.20/28,665.10 = 0.0238$.

EXAMPLE 12.9—Test the significance of the differences among adjusted block yields, proceeding exactly as in table 12.14 with this exception—make up the new sum by adding *block* and *error* figures. The F value is 11.4, highly significant. How do you interpret this result? Examine the adjusted block means. Note the negative correlation in table 12.14.

EXAMPLE 12.10—It was suggested that superphosphate might have a fungicidal action preventing the loss of the young plants, or that it might stimulate rapid growth

during the susceptible stage enabling the seedlings to escape infection. In the experimental results of table 12.13 is there any evidence about this question? It is easy to test the significance of the difference between adjusted means of the yields of plots with and without superphosphate. First, analyze the 6 degrees of freedom for treatment into two parts, the first having the one degree of freedom representing the contrast between plots with superphosphate and those without, while the other five are those within the groups of plots just contrasted. The sums and sums of squares and products of observations for the four plots with superphosphate and for the three lacking it are readily made up from table 12.13. You are then ready to compute this table:

Source of Variation	Degrees of Freedom	Sx^2	Sxy	Sy^2
Treatments (table 12.14)	6	116,020.33	3,598.05	112.8562
<i>P</i> vs. none	1	101,348.43	3,218.71	102.2221
Lots in the same group	5	14,671.90	379.34	10.6341

This table shows that a large part of the variation in both stand and yield is between the two groups having and lacking superphosphate. It is especially interesting to test the significance of the difference between the adjusted means of these two groups:

Source of Variation	Sx^2	Sxy	Sy^2	Errors of Estimate		
				Sum of Squares	Degrees of Freedom	Mean Square
<i>P</i> vs. none	101,348.43	3,218.71	102.2221		1	0.241
Error (table 12.14)	28,665.10	682.20	23.2326	6.9969	29	0.241
Sum	130,013.53	3,900.91	125.4547	8.4123	30	
Difference for testing significance				1.4154	1	1.415*

Adjusted mean yields: *P*, 5.626; None, 4.834

This shows that the superphosphate produced a significantly larger yield after adjustment for stand. This effect is small, however, compared with the differences in yield due to stand. In this comparison there is a defect that will be remedied in example 15.12. The conclusions are unaltered.

EXAMPLE 12.11—The following data are adapted from those of Crampton (1) in such fashion that the means, the error correlation, and the tests of significance are little changed. *X* is food consumed, decreased by 300 pounds for each pig. *Y* is gain in weight, less 50 pounds. The pigs were assigned to the lots in groups of three, the three pigs in each group being litter mates of approximately equal weights.

Group	Lot					
	1		2		3	
	<i>X</i>	<i>Y</i>	<i>X</i>	<i>Y</i>	<i>X</i>	<i>Y</i>
1	84	17	77	38	127	38
2	16	7	103	45	73	29
3	60	34	60	30	57	35
4	89	37	52	46	48	40
5	21	10	13	21	80	23

By supplying expected values, verify the fact that the error correlation is 0.7312. Verify this also by analysis of covariance. Test the significance of group differences after adjustment for food consumed, $F = 4.66$, $d.f. = 4$ and 7; also of adjusted treatment differences, $F = 6.30$, $d.f. = 2$ and 7.

12.8—Covariance in a latin square. The foregoing covariance methods are readily applied when two variates are measured in a latin square experiment. For illustration we have taken the data in table 12.17 from the experiment on millet (2) partly analyzed in table 11.11. The additional data on stooling are now introduced to learn if the number of culms affects the yield. The sums of squares and products are listed in table 12.18. The only new feature is the presence of both rows and columns instead of only blocks. The data for the computation of corresponding sums of squares are taken from the column and row of sums, table 12.17. The sums of squares and products for spacings are calculated from the sums, SX and SY , in the appendix at the bottom of the table.

TABLE 12.17

MEAN NUMBER OF CULMS PER PLANT, X , AND YIELD (GRAMS), Y , OF MILLET IN THE PLOTS OF A LATIN SQUARE

The five spacings are denoted by (2), (4), (6), (8), and (10) inches.

Row	No. X or Yield Y	Column					Sums SX SY
		1	2	3	4	5	
1	X Y	(4) 3.77 257	(10) 4.02 230	(2) 3.50 279	(6) 3.97 287	(8) 3.60 202	18.86 1255
2	X Y	(8) 4.03 245	(2) 3.00 283	(10) 5.37 245	(4) 4.02 280	(6) 3.60 260	20.02 1313
3	X Y	(10) 4.02 182	(4) 3.00 252	(6) 3.02 280	(8) 3.73 246	(2) 2.90 250	16.67 1210
4	X Y	(2) 2.60 203	(6) 4.33 204	(8) 3.30 227	(10) 3.33 193	(4) 3.33 259	16.89 1086
5	X Y	(6) 3.83 231	(8) 4.03 271	(4) 4.00 266	(2) 2.47 334	(10) 5.53 338	19.86 1440
Sum	SX SY	18.25 1118	18.38 1240	19.19 1297	17.52 1340	18.96 1309	92.30 6304

$$SX^2 = 353.3144$$

$$SXY = 23,326.36$$

$$SY^2 = 1,626,188$$

		Spacing Sums and Means				
		(2)	(4)	(6)	(8)	(10)
Sums	SX SY	14.47 1,349	18.12 1,314	18.75 1,262	18.69 1,191	22.27 1,188
Means	X Y	2.89 269.8	3.62 262.8	3.75 252.4	3.74 238.2	4.45 237.6

TABLE 12.18
COVARIANCE IN A LATIN SQUARE EXPERIMENT ON THE EFFECTS OF SPACING UPON THE
STOOLING AND YIELD OF MILLET

Source of Variation	Degrees of Freedom	Sums of Squares and Products			Errors of Estimate		
		Sx^2	Sxy	Sy^2	Sum of Squares	Degrees of Freedom	Mean Square
Total	24	12.5428	51.99	36,571.36			
Rows	4	2.0445	139.07	13,601.36			
Columns	4	0.3434	1.54	6,146.16			
Spacings	4	6.1378	-132.62	4,156.56			
Error	12	4.0171	44.00	12,667.28	12,185.34	11	1,107.8
Spacing plus error	16	10.1549	- 88.62	16,823.84	16,050.47	15	
Difference for testing significance of spacing					3,865.13	4	966.3

Since we are interested only in the effect of spacing on yield per stool, we make the test of significance at the bottom of table 12.18. The adjusted means vary actually less than one might expect in 50% of trials from a homogeneous population. It makes no difference, then, whether the adjusted or unadjusted yields are tested, the differences are not significant.

You may raise a pertinent question at this point if you haven't done so before. Why should yield be adjusted for number of culms? If close planting were found to give significantly greater yields, what difference would it make if the number of culms were less? The physiologist and the producer might differ in their answers to the question. Whether to adjust for number of culms or not depends upon the information desired. It is for the investigator to decide what question he is proposing. Don't use covariance unless there is a well-considered reason for doing so.

EXAMPLE 12.12—Test the significance of the differences among row yields of millet. You will get $F = 3.22$, $df. = 4$ and 12. Then test the significance of differences among adjusted row yields. Your test will be based on $F = 1.93$, $df. = 4$ and 11.

EXAMPLE 12.13—In the manner of example 11.26, insert in table 12.17 expected values for both culms and yields. Make a table of 25 pairs of deviations from expected. Calculate r , then verify by use of the regression statistics in the line for error, table 12.18.

REFERENCES

1. EARLE WILCOX CRAMPTON. *Journal of Nutrition*, 7:305 (1934).
2. H. W. LI, C. J. MENG, and T. N. LIU. *Journal of the American Society of Agronomy*, 28:1 (1936).
3. GEORGE W. SNEDECOR and C. C. CULBERTSON. *Record of Proceedings, American Society of Animal Production* (1932).

REFERENCES NOT CITED

- M. S. BARTLETT. *Journal of Agricultural Science*, 25:238 (1935).
- R. A. FISHER. *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh, Section 49.1 (1934).
- E. S. PEARSON and S. S. WILKS. *Biometrika*, 25:353 (1933).
- GEORGE W. SNEDECOR and GERTRUDE M. COX. *Journal of Agricultural Research*, 54:449 (1937).
- HARRY W. TITUS and H. M. HARSHAW. *Poultry Science*, 14:3 (1935).
- B. L. WELCH. *Biometrika*, 27:145 (1935).
- JOHN WISHART and H. G. SANDERS. *Principles and Practice of Field Experimentation*. The Empire Cotton Growing Corporation, London (1935).

Chapter 13

MULTIPLE REGRESSION AND COVARIANCE

13.1—Introduction. Drawing together the threads of the foregoing methods, multiple regression completes a fabric of increasingly complex patterns. Starting with a unique sample, the number of variates and the number of groups have been alternately expanded. We are now ready to make another extension, increasing the number of variates beyond two. Since three or more characters are often measured, their mutual relationships must be evaluated. At present three variates in a single group will be considered. Afterwards multiple groups will receive attention. Four and more variates offer no new principles, only increasing the extent and intricacy of the calculations.

13.2—Three variates in a single group. The increase from two variates to three adds a surprising number of new concepts to those developed in chapters 6 and 7. Perhaps the most striking is the necessity for three dimensions in which to depict the variates graphically. Solid geometry is required instead of plane. If Y partly depends upon X_1 and X_2 for its value, three mutually perpendicular axes are demanded for graphical presentation. X_1 is measured along one, X_2 is laid out parallel to the second and Y rises from the plane X_1X_2 parallel to the third axis. The points in space fixed by the triplets of values (X_1, X_2, Y) determine a regression plane whose equation is

$$\hat{Y} = b + b_{Y1.2} X_1 + b_{Y2.1} X_2,$$

$b_{Y1.2}$ and $b_{Y2.1}$ being the *partial regression coefficients*. The first is read, "the regression of Y on X_1 independent of X_2 ."

This leads to a second complication. Among the three variates there are three correlation coefficients, r_{Y1} , r_{Y2} , and r_{12} .† How may these be combined into two regression coefficients? The principle of least squares is again invoked. The regression plane is designed to pass among the points so that the sum of squares of vertical distances from plane to points shall be a minimum—the least, that is, for any plane surface. Curvilinear regression is not considered in this chapter.

The methods will be illustrated by use of the data on pigs, table 13.1. These animals were kept on adequate rations during the period of approximately uniform growth rate. We wish to evaluate the relationships

† The *correlation method* will be described first. Beginning with section 13.11, the *alternative direct method* will be presented.

among initial weight, initial age and average daily gain, and to inquire the amount of information about rate of gain which is furnished in advance by the two independent variates.

The computations are conveniently arranged as in table 13.2. The sums of the 40 items, the sums of their squares and of their products (lines 1) are entered in the table from the dials of the calculating machine. In lines 2 are the usual correction terms. As illustrations, in line X_12 ,

TABLE 13.1
INITIAL AGE (X_1), INITIAL WEIGHT (X_2), AND RATE OF GAIN (Y) OF 40 SWINE
GROWING UNIFORMLY

Pig No.	Initial Age X_1 (days)	Weight X_2 (pounds)	Rate of Gain Y (pounds per day)	Estimated Rate $\hat{Y}$	Error of Estimate $Y - \hat{Y}$
1	78	61	1.40	1.45	-.05
2	90	59	1.79	1.39	.40
3	94	76	1.72	1.50	.22
4	71	50	1.47	1.39	.08
5	99	61	1.26	1.38	-.12
6	80	54	1.28	1.39	-.11
7	83	57	1.34	1.40	-.06
8	75	45	1.55	1.34	.21
9	62	41	1.57	1.35	.22
10	67	40	1.26	1.33	-.07
11	78	74	1.61	1.54	.07
12	99	75	1.31	1.48	-.17
13	80	64	1.12	1.46	-.34
14	75	48	1.35	1.36	-.01
15	94	62	1.29	1.40	-.11
16	91	42	1.24	1.26	-.02
17	75	52	1.29	1.39	-.10
18	63	43	1.43	1.36	.07
19	62	50	1.29	1.42	-.13
20	67	40	1.26	1.33	-.07
21	78	80	1.67	1.59	.08
22	83	61	1.41	1.43	-.02
23	79	62	1.73	1.45	.28
24	70	47	1.23	1.37	-.14
25	85	59	1.49	1.41	.08
26	83	42	1.22	1.29	-.07
27	71	47	1.39	1.37	.02
28	66	42	1.39	1.35	.04
29	67	40	1.56	1.33	.23
30	67	40	1.36	1.33	.03
31	77	62	1.40	1.46	-.06
32	71	55	1.47	1.42	.05
33	78	62	1.37	1.45	-.08
34	70	43	1.15	1.34	-.19
35	95	57	1.22	1.36	-.14
36	96	51	1.48	1.31	.17
37	71	41	1.31	1.32	-.01
38	63	40	1.27	1.34	-.07
39	62	45	1.22	1.38	-.16
40	67	39	1.36	1.32	.04

TABLE 13.2
CALCULATION OF SUMS OF SQUARES AND PRODUCTS AND CORRELATION
COEFFICIENTS FOR PIG DATA

$n = 40$		X_1	X_2	Y
Sums		3,082	2,109	55.53
Means		77.050	52.725	1.3882
X_1	1. SX_1^2 , SX_1X_2 , SX_1Y	242,204	165,536	4,285.51
	2. Correction terms	237,468.10	162,498.45	4,278.5865
	3. Sx_1^2 , Sx_1x_2 , Sx_1y	4,735.90	3,037.55	6.9235
	4. $\sqrt{Sx_1^2}$, $\sqrt{(Sx_1^2)(Sx_2^2)}$, etc.	68.8179	4,898.16	69.5983
	5. r_{12} , r_{Y1}		0.6201	0.0995
X_2	1. SX_2^2 , SX_2Y		116,263	2,955.36
	2. Correction terms		111,197.02	2,927.8192
	3. Sx_2^2 , Sx_2y		5,065.98	27.5408
	4. $\sqrt{Sx_2^2}$, $\sqrt{(Sx_2^2)(Sy^2)}$		71.1757	71.9828
	5. r_{Y2}			0.3826
Y	1. SY^2			78.1123
	2. Correction term			77.0895
	3. Sy^2			1.0228
	4. $\sqrt{Sy^2}$			1.01134
	5. s_y			0.162

column X_1 , $(3,082)^2/40 = 237,468.10$; while in line X_22 , column Y , $(2,109)(55.53)/40 = 2,927.8192$. The differences in lines 3 are sums of squares and products of deviations from means. In each line 4 the first number is the square root of the sum of squares just above. The other numbers in these lines are products of these square roots. For example, in line X_14 , column Y , $\sqrt{(Sx_1^2)(Sy^2)} = (68.8179)(1.01134) = 69.5983$. These products are the denominators of the correlation coefficients, the coefficients being calculated in the usual manner; as an example:

$$r_{12} = \frac{3,037.55}{4,898.16} = 0.6201$$

13.3—The multiple regression equation. Now comes the first new feature in the calculation of multiple regression. The application of the principle of least squares leads to a pair of simultaneous *normal* equations,

$$\begin{aligned} b'_{Y1.2} + r_{12} b'_{Y2.1} &= r_{Y1} \\ r_{12} b'_{Y1.2} + b'_{Y2.1} &= r_{Y2} \end{aligned}$$

in which $b'_{Y1.2}$ and $b'_{Y2.1}$ represent the *standard partial regression coefficients* that will be used to compute the regression of Y on X_1 and X_2 . The *b*-primes are used in order to distinguish these standard regression co-

efficients from the b 's of the equation in the preceding section. The relation between the b -primes and the b 's will emerge presently. The symbol $b'_{Y1.2}$ indicates the *standard regression of Y on X_1 independent of X_2* .

The solution of the normal equations yields the formulas,

$$b'_{Y1.2} = \frac{r_{Y1} - r_{Y2} r_{12}}{1 - r_{12}^2}, \quad b'_{Y2.1} = \frac{r_{Y2} - r_{Y1} r_{12}}{1 - r_{12}^2}$$

Substituting the pig data,

$$b'_{Y1.2} = \frac{0.0995 - (0.3826)(0.6201)}{1 - (0.6201)^2} = -0.2239$$

$$b'_{Y2.1} = \frac{0.3826 - (0.0995)(0.6201)}{1 - (0.6201)^2} = 0.5214$$

If we use standard measure (section 7.4), the regression equation is

$$\begin{aligned} \hat{y} &= b'_{Y1.2} x_1' + b'_{Y2.1} x_2' \\ &= -0.2239x_1' + 0.5214x_2', \end{aligned}$$

where $x_1' = (X_1 - \bar{x}_1)/s_1$, etc. That is, the b -primes are the regression coefficients if standard measure is used. Since the ordinary units of measurements must be returned to, the equation is usually given in some such form as this:

$$\hat{Y} = \bar{y} + b'_{Y1.2}(s_y/s_1)(X_1 - \bar{x}_1) + b'_{Y2.1}(s_y/s_2)(X_2 - \bar{x}_2)$$

It is clear that the b 's are computed from the b -primes through multiplication by the ratio of two standard deviations containing the desired units of measurements. Usually it is more convenient to calculate the equation,

$$\hat{Y} = \bar{y} + b'_{Y1.2} \frac{\sqrt{S_y^2}}{\sqrt{S_{x_1}^2}}(X_1 - \bar{x}_1) + b'_{Y2.1} \frac{\sqrt{S_y^2}}{\sqrt{S_{x_2}^2}}(X_2 - \bar{x}_2)$$

Substituting the b -primes and data from table 13.2,

$$\begin{aligned} \hat{Y} &= 1.3882 - 0.2239 \frac{1.01134}{68.8179}(X_1 - 77.05) \\ &\quad + 0.5214 \frac{1.01134}{71.1757}(X_2 - 52.725) \\ &= 1.3882 - 0.003290(X_1 - 77.05) + 0.007409(X_2 - 52.725) \\ &= 1.3882 - 0.003290X_1 + 0.2535 + 0.007409X_2 - 0.3906 \\ &= 1.251 - 0.003290X_1 + 0.007409X_2 \end{aligned}$$

The regression coefficients indicate that in this sample of pig data the average daily gain decreased 0.003290 pound per day with each day

increase in initial age, but increased 0.007409 pound per day with each pound increase of initial weight. Both coefficients specify changes in Y quite independent of changes in the other independent variate. However, one should not forget that age and weight tend to increase together ($r_{12} = 0.6201$); hence, in changing from one individual to another, Y usually gets both a positive and negative increment with a tendency to partially counterbalance. To see this, consider two pigs whose errors of estimate are small:

Number	Age	Weight	Average Daily Gain
14	75	48	1.35
22	83	61	1.41

Since the second is 8 days older than the first, its rate of gain is expected to be $(8)(0.003290) = 0.0263$ pound less; but since it weighs 13 pounds more its rate is expected to be $(13)(0.007409) = 0.0963$ greater. These expected changes being concurrent, the net difference in rate should be $0.0963 - 0.0263 = 0.07$ pound to be compared with actual difference, $1.41 - 1.35 = 0.06$ pound.

You will be interested in the positive correlation between age and rate, $r_{Y1} = 0.0995$, and the negative regression, $b_{Y1.2} = -0.003290$. At first this is puzzling. In the sample as a whole gain increases slightly with initial age, but in groups having the same initial weight gains are smaller among the older pigs. It is the same kind of relation you might have found in the foregoing chapter if the total regression had been positive when that within groups was negative. An easy way to clarify thinking is to divide the pigs into weight classes, as in table 13.3, then treat them in the manner of covariance. If weight is ignored, the *total* correlation is $r_{Y1} = 0.0995$, with total regression,

$$\hat{Y} = 1.388 + 0.0995 \frac{1.0113}{68.82}(X_1 - 77.05) = 0.001462X_1 + 1.275$$

Within the weight classes, however, the regressions are predominantly negative (table 13.4), the average being -0.003347 . (The small discrepancy between this value and that of the partial regression coefficient, -0.003290 , is due to the slightly different hypotheses involved.) For your thinking, then, you have the average regression of gain on age within weight classes, -0.003347 , contrasted with the overall regression, 0.001462 . A graph similar to that of figure 12.2 will clarify the distinction. Putting all these ideas together, you will see readily that the regression of gain on age within the initial weight groups is the same kind of relation indicated by $b_{Y1.2}$. It is independent of X_2 because all the pigs of a group have approximately the same weight: it is an estimate of the regression in a population of pigs all starting with the same weight. In a similar

TABLE 13.3
DATA ON 40 SWINE (TABLE 13.1) CLASSIFIED BY INITIAL WEIGHT

Initial Weight	Number of Swine	Age and Average Daily Gain	Total	Mean
39-44	13	62 67 91 63 67 83 66 67 1.57 1.26 1.24 1.43 1.26 1.22 1.39 1.56		
		67 70 71 63 67 1.36 1.15 1.31 1.27 1.36	904 17.38	69.5 1.34
45-49	5	75 75 70 71 62 1.55 1.35 1.23 1.39 1.22	353 6.74	70.6 1.35
50-54	5	71 80 75 62 96 1.47 1.28 1.29 1.29 1.48	384 6.81	76.8 1.36
55-59	5	90 83 85 71 95 1.79 1.34 1.49 1.47 1.22	424 7.31	84.8 1.46
60-64	8	78 99 80 94 83 79 77 78 1.40 1.26 1.12 1.29 1.41 1.73 1.40 1.37	668 10.98	83.5 1.37
74-80	4	78 94 99 78 1.61 1.72 1.31 1.67	349 6.31	87.2 1.58
Total	40		3,082 55.53	77.05 1.388

manner, if the pigs were divided into groups having approximately the same initial age the average regression of gain on initial weight, independent of age, would be found to be about 0.007409 pound per pound of initial weight. By use of multiple regression we accomplish in a single equation the segregation of the independent contributions of X_1 and X_2 to the dependent variate Y .

If you wish to get an idea of the relative potency of age and weight in estimating rate of gain, you must avoid comparing the b -coefficients.

TABLE 13.4
ANALYSIS OF COVARIANCE IN WEIGHT CLASSES OF SWINE

Weight Class	Degrees of Freedom	Sums of Squares and Products			Regression of $\bar{Y}$ on X_1
		Sx_1^2	Sx_1y	Sy^2	
39-44	12	831.2308	-6.1885	0.1917	-0.007445
45-49	4	113.2000	2.0860	0.0729	0.018428
50-54	4	634.8000	2.5720	0.0427	0.004052
55-59	4	324.8000	-0.6480	0.1819	-0.001995
60-64	7	486.0000	-3.6700	0.2140	-0.007551
74-80	3	354.7500	-3.3375	0.1015	-0.009408
Sum	34	2,744.7808	-9.1860	0.8047	-0.003347

They are fractions with s_1 in one denominator and s_2 in the other. If these two standard deviations were very different the two regression coefficients would reflect this consequence of units of measurement. It is the b' -coefficients that should be compared. Look back at the regression equation in standard measure. If X_1 and X_2 are normally distributed, then x_1' and x_2' are both estimates of normal deviates with the same range and with standard deviations equal to 1. Changes in x_1' and x_2' are reflected in $\hat{y}$ according to the values of $b'_{Y1.2} = -0.2239$ and $b'_{Y2.1} = 0.5214$, respectively. Naturally, you can't be too sure of population parameters from a comparison of these two statistics. They are subject to sampling variation. But in the sample itself the b -primes give the information desired. Among the data of table 13.1, initial weight is more than twice as effective in forecasting as is initial age.

EXAMPLE 13.1—Here is a set of 20 triplets arranged for easy computation:

X_1	X_2	Y	X_1	X_2	Y
29	2	22	21	2	24
1	4	26	12	1	7
5	3	23	24	3	23
31	1	8	16	3	28
25	3	25	6	4	25
16	1	12	20	2	22
26	1	13	35	1	25
15	4	30	9	4	32
6	2	12	19	4	37
10	3	26	14	2	20

Calculate the regression,

$$\hat{Y} = 0.35X_1 + 7.16X_2 - 1.9$$

EXAMPLE 13.2—In the foregoing example compare $r_{Y1} = -0.1354$ with $b_{Y1.2} = 0.3538$. Make a table like 13.3 classifying the data in four groups having X_2 equal respectively to 1, 2, 3, and 4. Except for the values of X_2 the table will be the same as example 12.1. Calculate the average regression of Y on X_1 within the groups. You will get 0.3621 as a companion for $b_{Y1.2} = 0.3538$.

EXAMPLE 13.3—Compute the results recorded in table 13.4.

13.4—Errors of estimate. The uses of the multiple regression equation are much the same as those described for the two-variable regression in chapter 6. One may estimate the rate of gain of each animal if he wishes. Pig number 1 is expected to gain

$$\hat{Y} = 1.251 - (0.003290)(78) + (0.007409)(61) = 1.45 \text{ pounds per day}$$

Actually the gain was 1.40 pounds per day, the error of estimate being $1.40 - 1.45 = -0.05$ pound per day. These results are entered in the last two columns of table 13.1. In accordance with the theory, the sum of the errors of estimate is zero, while the sum of their squares is the least possible for any linear regression. (Geometrically, *linear regression* in three dimensions represents a plane.)

While one isn't ordinarily much interested in the errors of estimate

for individual pigs, there are experiments where these data are the chief objective of the investigation. Let us see what kinds of information they contain. The one largest in absolute value is that for pig number 2. Being 3 months old with a weight of only 59 pounds, he was expected to gain only 1.39 pounds per day, but actually made the best gain of all, 1.79 pounds per day. One might be much concerned to know why. One might also wish to compare the performance of this pig with that of number 23. If no attention is paid to initial age and weight, the difference in rate of gain is only $1.79 - 1.73 = 0.06$ pound per day. But that comparison would be unfair to number 2 when X_1 and X_2 are taken into account. For his age and weight, number 2 was 0.40 pound per day better than average, whereas number 23 was only 0.28 better, a difference in favor of number 2 of 0.12 pound per day. Multiple regression enables one to evaluate individuals and to compare them on the basis of their deviations from average performance, due account being taken of known information contained in the independent variates.

Recalling the facts of section 7.7, one may wish to know the correlation between Y and $\hat{Y}$ in table 13.1. It could be calculated in the manner of chapter 7, but a short cut is available. The result is known as the *multiple correlation coefficient*, and is arrived at by this formula:

$$\begin{aligned} R^2 &= r_{Y1} b'_{Y1.2} + r_{Y2} b'_{Y2.1} \\ &= (0.0995)(-0.2239) + (0.3826)(0.5214) = 0.1772, \end{aligned}$$

whence $R = 0.421$ is the correlation one would get by the routine calculation. This coefficient measures the success of estimating rate of gain from the two preliminary items of information. It is always less than unity, but greater than either of the coefficients r_{Y1} or r_{Y2} , a fact which is sometimes valuable in detecting errors in calculation.

The variation in the estimated values, $\hat{Y}$, of table 13.1 is due to regression: if $\hat{Y}$ were plotted against X_1 and X_2 in a three-dimensional graph, every plotted point would lie in the regression plane. On the contrary, the errors of estimate result from failure in estimating. In the graph they would be represented by line segments extending above or below the regression plane to the points represented by (X_1, X_2, Y) . In table 13.1 one could run up the sum of squares of deviations from mean in each of these last two columns. He would find them to be respectively (see section 7.8 for an analogy):

$$\begin{aligned} R^2 S_y^2 &= (0.1772)(1.0228) = 0.1812 \\ (1 - R^2) S_y^2 &= (1 - 0.1772)(1.0228) = 0.8416 \end{aligned}$$

That is, the total sum of squares is separated into two parts, the first, 0.1812, explained by regression, and the second, 0.8416, independent of regression. The latter is the sum of squares of errors of estimate, leading directly to the *standard error of estimate*,

$$s_{Y.12} = \sqrt{\frac{(1 - R^2) S_y^2}{n - 3}} = \sqrt{\frac{(1 - 0.1772)(1.0228)}{37}} = 0.151 \text{ pound per day}$$

This measures variation among errors of estimate in the same way that s_y measures variation in average daily gains. The introduction of the two control variates has resulted in a reduction of the unexplained variation from $s_y = 0.162$ (table 13.2) to $s_{Y.12} = 0.151$ pound per day. The degrees of freedom, $n - 3$, are less than n by the number of variates. Actually, three independent averages have been used in the regression, the mean of Y and the two partial regression coefficients. One degree of freedom is deducted for each.

If the portions of information furnished by X_1 and X_2 tend to be independent, r_{12} is likely small and R may be much larger than either r_{Y1} or r_{Y2} . In the pig data the facts are otherwise. X_1 and X_2 are correlated rather strongly, age and weight being concurrent phenomena in young growing animals. It is not surprising, therefore, to find $R = 0.421$ little greater than $r_{Y2} = 0.3826$. Almost as good a job of estimating can be done with weight alone as independent variate. Age contributes little additional information.

Occasionally one is not so much interested in the partitioning of Sy^2 into R^2Sy^2 and $(1 - R^2)Sy^2$ as in a corresponding division of the variance, $s_y^2 = Sy^2/(n - 1)$. For any number of variates, m , this is done in table 13.5. The mean square for deviations is expressed in terms of s^2 by substituting for Sy^2 its equal, $(n - 1)s^2$, giving the exact variance in the last line of the table. Subtraction from the total variance, s^2 , leaves the variance attributable to regression in the second line.

TABLE 13.5
ANALYSIS OF VARIANCE IN REGRESSION; m VARIATES

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	Variance	
				Exact	n Large
Total	$n - 1$	Sy^2	$\frac{Sy^2}{n - 1}$		s^2
Regression	$m - 1$	R^2Sy^2	$\frac{R^2Sy^2}{m - 1}$	$\left[1 - (1 - R^2)\frac{n - 1}{n - m}\right]s^2$	R^2s^2
Deviations	$n - m$	$(1 - R^2)Sy^2$	$\frac{(1 - R^2)Sy^2}{n - m}$	$\left[(1 - R^2)\frac{n - 1}{n - m}\right]s^2$	$(1 - R^2)s^2$

tuting for Sy^2 its equal, $(n - 1)s^2$, giving the exact variance in the last line of the table. Subtraction from the total variance, s^2 , leaves the variance attributable to regression in the second line.

An adjusted value of R^2 ,

$$R_A^2 = 1 - (1 - R^2)\frac{n - 1}{n - m},$$

simplifies the formulas for exact variance. In terms of R_A^2 the total variance, s^2 , is seen to be partitioned into the two fractions,

$$R_A^2s^2, \text{ attributed to regression, and} \\ (1 - R_A^2)s^2, \text{ deviations from regression}$$

In large samples the fraction, $(n-1)/(n-m)$, differs little from 1; hence, R_A^2 is approximately equal to R^2 . This is the origin of the not uncommon but approximate statement that the variance of the dependent variate is the sum of the two parts, R^2s^2 due to regression and $(1-R^2)s^2$, deviations from regression.

EXAMPLE 13.4—Calculate $R = 0.8837$ for the data of example 13.1.

EXAMPLE 13.5—Compute the regression of X_1 on Y and X_2 in example 13.1:

$$\hat{X}_1 = 1.0289Y - 10.162X_2 + 19.77, \quad R = 0.7396$$

Save your work sheets for use later.

EXAMPLE 13.6—Derive the two formulas:

$$R^2 = \frac{r_{Y1}^2 + r_{Y2}^2 - 2r_{Y1}r_{Y2}r_{12}}{1 - r_{12}^2}$$

$$R^2 = b'_{Y1.2}^2 + b'_{Y2.1}^2 + 2r_{12}b'_{Y1.2}b'_{Y2.1}$$

EXAMPLE 13.7—Do you think it would be possible to get $r_{Y1} = 0.6$, $r_{Y2} = 0.8$, and $r_{12} = -0.5$ from a set of experimental data?

EXAMPLE 13.8—Starting with $r_{Y1} = 0.6$ and $r_{Y2} = 0.8$, try the effects on R produced by various values of r_{12} .

EXAMPLE 13.9—Calculate $R = 0.89$ starting with $r_{Y1} = 0.21$, $r_{Y2} = 0.19$, and $r_{12} = -0.9$. Think of X_1 as the temperature in a greenhouse, X_2 , the relative humidity and Y the rate of growth of a plant.

13.5—Tests of significance. All of the statistics of multiple regression are subject to sampling variation, of course, so that measures of variation and tests of significance are in order. The standard errors of the two b -primes are identical,

$$\sqrt{\frac{1 - R^2}{(1 - r_{12}^2)(n-3)}} = \sqrt{\frac{1 - 0.1772}{[1 - (0.6201)^2](37)}} = 0.190$$

Significance is tested by means of t ; that is, the usual null hypothesis is set up and the probability of exceeding the sample value is determined. In our illustrative example,

$$\begin{aligned} \text{for } b'_{Y1.2}: t &= 0.2239/0.190 = 1.18, \\ \text{while for } b'_{Y2.1}: t &= 0.5214/0.190 = 2.74^{**} \end{aligned}$$

The signs of the b -primes are ignored. For 37 *df.* the second is highly significant.

Irrespective of the results from testing the b -primes, the significance of the multiple regression as an entity may be tested. The test, identical with the test of the significance of R , is effected by using F to determine whether the mean square due to regression is significantly greater than

that unexplained by regression. The test is conveniently thrown into the form of analysis of variance:

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Regression, $R^2 S_y^2$	2	0.1812	0.0906
Errors of estimate, $(1 - R^2) S_y^2$	37	0.8416	0.0227
Total	39	1.0228	

$$F = 0.0906/0.0227 = 3.99^*$$

For a more convenient test table 13.6 is entered for R (10) in the same manner as table 7.3 was for r . One merely has to look at the left for the degrees of freedom, $n - 3$, then read in the column for 3 variables the values of R at the 5% and 1% points. If the sample value is less than the 5% point, evidence against the null hypothesis is considered to be too slight for significance. Values of R between the two points are called significant, while those greater than the 1% point are designated highly significant. In the illustrative example, $R = 0.421$, $d.f. = 37$, lies between the 5% and 1% points. Such a value is not likely to arise in sampling from a population with zero R , but occasionally may.

Naturally the results from the two methods which have been given for testing R must be identical. Use the one that is more available. For convenience, the 5% and 1% values of r copied from table 7.3, are entered in the column for 2 variables, table 13.6. The other two columns will be explained later.

As in chapter 6, one can now make fiducial statements about the mean, $\bar{y}$, about the regression coefficients and about $\hat{Y}$ or Y . For convenience, this topic will be deferred to section 13.11.

13.6—Multiple covariance in groups. Experimental data aren't often taken in single groups like that of table 13.1. Animals are too expensive, time is too precious and too many treatments are awaiting trial. Usually some design like those in chapters 10 and 11 is used. These are all readily adapted to the necessities of three variables. Two examples will be worked out.

As was mentioned earlier, the 40 swine of table 13.1 really constituted four lots whose rations and gains were not greatly different. We shall illustrate the simplest case of multiple covariance (more than two variates in two or more groups) by means of the four lots of pig data. We wish to test the significance of the differences among adjusted means in the manner of section 12.3. Since there are now three variates instead of two, the methods of multiple regression are required.

If we were starting with a new sample of data we should first calculate the multiple regression of the entire set exactly as has been done in the

TABLE 13.6
THE 5% AND 1% POINTS FOR t AND R

Degrees of Freedom	Number of Variables				Degrees of Freedom	Number of Variables			
	2	3	4	5		2	3	4	5
1	.997 1.000	.999 1.000	.999 1.000	.999 1.000	24	.388 .496	.470 .565	.523 .609	.562 .642
2	.950 .990	.975 .995	.983 .997	.987 .998	25	.381 .487	.462 .555	.514 .600	.553 .633
3	.878 .959	.930 .976	.950 .983	.961 .987	26	.374 .478	.454 .546	.506 .590	.545 .624
4	.811 .917	.881 .949	.912 .962	.930 .970	27	.367 .470	.446 .538	.498 .582	.536 .615
5	.754 .874	.836 .917	.874 .937	.898 .949	28	.361 .463	.439 .530	.490 .573	.529 .606
6	.707 .834	.795 .886	.839 .911	.867 .927	29	.355 .456	.432 .522	.482 .565	.521 .598
7	.666 .798	.758 .855	.807 .885	.838 .904	30	.349 .449	.426 .514	.476 .558	.514 .591
8	.632 .765	.726 .827	.777 .860	.811 .882	35	.325 .418	.397 .481	.445 .523	.482 .556
9	.602 .735	.697 .800	.750 .836	.786 .861	40	.304 .393	.373 .454	.419 .494	.455 .526
10	.576 .708	.671 .776	.726 .814	.763 .840	45	.288 .372	.353 .430	.397 .470	.432 .501
11	.553 .684	.648 .753	.703 .793	.741 .821	50	.273 .354	.336 .410	.379 .449	.412 .479
12	.532 .661	.627 .732	.683 .773	.722 .802	60	.250 .325	.308 .377	.348 .414	.380 .442
13	.514 .641	.608 .712	.664 .755	.703 .785	70	.232 .302	.286 .351	.324 .386	.354 .413
14	.497 .623	.590 .694	.646 .737	.686 .768	80	.217 .283	.269 .330	.304 .362	.332 .389
15	.482 .606	.574 .677	.630 .721	.670 .752	90	.205 .267	.254 .312	.288 .343	.315 .368
16	.468 .590	.559 .662	.615 .706	.655 .738	100	.195 .254	.241 .297	.274 .327	.300 .351
17	.456 .575	.545 .647	.601 .691	.641 .724	125	.174 .228	.216 .266	.246 .294	.269 .316
18	.444 .561	.532 .633	.587 .678	.628 .710	150	.159 .208	.198 .244	.225 .270	.247 .290
19	.433 .549	.520 .620	.575 .665	.615 .698	200	.138 .181	.172 .212	.196 .234	.215 .253
20	.423 .537	.509 .608	.563 .652	.604 .685	300	.113 .148	.141 .174	.160 .192	.176 .208
21	.413 .526	.498 .596	.552 .641	.592 .674	400	.098 .128	.122 .151	.139 .167	.153 .180
22	.404 .515	.488 .585	.542 .630	.582 .663	500	.088 .115	.109 .135	.124 .150	.137 .162
23	.396 .505	.479 .574	.532 .619	.572 .652	1000	.062 .081	.077 .096	.088 .106	.097 .115

first part of this chapter. The sums of squares and products, table 13.2, are entered as *total* in the first lines of table 13.8, while the sum of squares of errors of estimate, $(1 - R^2)S_y^2 = 0.8416$, is copied in the first line of table 13.9.

Next, in order to segregate the effects of lots, we make a table of the lot sums, table 13.7. From these we calculate sums of squares and products

TABLE 13.7
DATA FOR FOUR LOTS OF 10 SWINE EACH, TAKEN FROM TABLE 13.1

Lot	Pig Numbers	Sums		
		Initial Age	Initial Weight	Average Daily Gain
1	1-10	799	544	14.64
2	11-20	784	550	13.19
3	21-30	749	520	14.45
4	31-40	750	495	13.25

for lots in the same way that table 13.2 was calculated for the individuals. The only difference is the use of the scheme of calculation already familiar in analysis of variance and covariance. As examples, the first entry^a in line 2 of table 13.8 is

$$\frac{(799)^2 + (784)^2 + (749)^2 + (750)^2}{10} - 237,468.10$$

$$= 237,665.80 - 237,468.10 = 187.70,$$

while the second, column X_2 , is

$$\frac{(799)(544) + \dots + (750)(495)}{10} - 162,498.45 = 160.15$$

The correction terms for lines 2 are the same as those of table 13.2.

The differences in lines 3 are the pooled sums of squares and products for individuals within lots, the first objective in this table. Two remarks are in order. First, the sums of squares and products might have been calculated directly for each lot, then pooled in the manner of table 12.6, but the use of lot sums is easier. Second, when there are only two subdivisions of the total, as in the present example, the correction terms need not be subtracted in lines 1 and 2. For example, Sx_1^2 in line 3 might well have been calculated from the uncorrected sums of squares,

$$242,204 - 237,655.80 = 4,548.20$$

We have followed the usual routine because of its familiarity and generality.

TABLE 13.8
SUMS OF SQUARES AND PRODUCTS FOR TOTAL, FOR LOTS, AND WITHIN LOTS. CALCULATION
OF CORRELATION COEFFICIENTS WITHIN LOTS

Source of Variation		Age $\bar{X}_1$	Weight $\bar{X}_2$	Average Daily Gain $\bar{Y}$
X_1	Total from table 13.2, line 3	4,735.90	3,037.55	6.9235
	Lots	187.70	160.15	1.3005
	Pigs within lots	4,548.20	2,877.40	5.6230
	$\sqrt{Sx_1^2}$, $\sqrt{(Sx_1^2)(Sx_2^2)}$, etc. r_{12} , r_{Y1}	67.4403	4,709.68 0.6110	62.0011 0.0907
X_2	Total from table 13.2, line 3		5,065.98	27.5408
	Lots		189.08	1.3218
	Pigs within lots		4,876.90	26.2190
	$\sqrt{Sx_2^2}$, $\sqrt{(Sx_2^2)(Sy^2)}$ r_{Y2}		69.8348	64.2025 0.4084
Y	Total from table 13.2, line 3			1.0228
	Lots			0.1776
	Pigs within lots $\sqrt{Sy^2}$			0.8452 0.9193

The second objective of table 13.8 is the computation of the three r 's from the pooled sums of squares and products within lots. These differ little from those of table 13.2 because the lot means are much alike.

From the three r 's *within lots* are now calculated $b'_{Y1.2} = -0.2535$, $b'_{Y2.1} = 0.5633$, and $R^2 = 0.20706$. Finally, the sum of squares of errors of estimate from this average regression within lots is

$$(1 - 0.20706)(0.8452) = 0.6702,$$

entered in line 2 of table 13.9. With the two sums of squares, total and error, the test of significance is made exactly as in table 12.3. The re-

TABLE 13.9
ANALYSIS OF VARIANCE OF AVERAGE DAILY GAINS AND OF ERRORS OF ESTIMATE
OF 40 SWINE IN 4 LOTS

Source of Variation	Degrees of Freedom	Ave. Daily Gain		R^2	Errors of Estimate		
		Sum of Squares	Mean Square		Degrees of Freedom	Sum of Squares	Mean Square
Total	39	1.0228		0.17722	37	0.8416	
Error	36	0.8452	0.0235	0.20706	34	0.6702	0.0197
Lots	3	0.1776	0.0592		3	0.1714	0.0571*

mainder, 0.1714 with $df. = 3$, is appropriate for testing the differences among lot mean rates of gain adjusted to average initial age and weight. The degrees of freedom in the first and second lines of the table are each less by two than the corresponding numbers available for analysis of variance. The deducted degrees of freedom correspond to the two partial regression coefficients used in these two lines.

Without information about initial weight and age, differences among lot means would be tested (table 13.9, column 3) by

$$F = 0.0592/0.0235 = 2.52,$$

somewhat short of the 5% point, 2.86. The reduction in error variance from 0.0235 to 0.0197 together with adjustment of the means to a common initial weight and gain have increased the precision of the experiment, and now

$$F = 0.0571/0.0197 = 2.90,$$

slightly beyond the 5% point.

EXAMPLE 13.10—Divide the data of table 13.1 into 2 parts, as follows:

Lots	Sum of Ages SX_1	Sum of Weights SX_2	Sum of Rates SY
1 and 3	1,548	1,064	29.09
2 and 4	1,534	1,045	26.44

Test the significance of the difference between the two adjusted mean rates of gain.
Ans. $F = 8.70$.

13.7—Multiple covariance with two criteria of classification.

Three or more variates may be required to give complete information about individuals classified by two criteria. A typical example is the randomized block in which one may have data on stand, percentage of plants with ears, and yield, for example, in varieties of corn. The variety yields may then be compared on the basis of equal stand and bearing plants. The computation is little different from that already described.

As illustration we select data from an interesting experiment (1) (11) designed to forecast the yield of wheat in Great Britain from measurements on the growing plant. In the present early stages of the investigation, it appears that most of the available information is contained in the two characters, shoot height at ear emergence (H) and plant numbers at tillering (N). Data for these variates together with yield are set out in table 13.10.

The problem is two-fold. First, a regression equation is needed for forecasting the year's yields. Second, the success in estimating must be gauged by applying the equation to the known data.

The desired equation must be free from the effects of season and

TABLE 13.10
 HEIGHTS OF SHOOTS AT EAR EMERGENCE (H), NUMBER OF PLANTS AT TILLERING (N),
 AND YIELD (Y) OF WHEAT IN GREAT BRITAIN
 (H , inches; N , number per foot; Y , cwt. per acre)

Year	Variate	Place						Year sums
		Seale Hayne	Rothamsted	Newport	Bog-hall	Sprows-ton	Plumpton	
1933	H	25.6	25.4	30.8	33.0	28.5	28.0*	171.3
	N	14.9	13.3	4.6	14.7	12.8	7.5	67.8
	Y	19.0	22.2	35.3	32.8	25.3	35.8*	170.4
1934	H	25.4	28.3	35.3	32.4	25.9	24.2	171.5
	N	7.2	9.5	6.8	9.7	9.2	7.5	49.9
	Y	32.4	32.2	43.7	35.7	28.3	35.2	207.5
1935	H	27.9	34.4	32.5	27.5	23.7	32.9	178.9
	N	18.6	22.2	10.0*	17.6	14.4	7.9	90.7
	Y	26.2	34.7	40.0	29.6	20.6	47.2	198.3
Place Sums	H	78.9	88.1	98.6	92.9	78.1	85.1	521.7
	N	40.7	45.0	21.4	42.0	36.4	22.9	208.4
	Y	77.6	89.1	119.0	98.1	74.2	118.2	576.2

* Estimated value.

locality. From the total sums of squares and products, therefore, must be deducted amounts based on year and place sums. The final stages of the computation are recorded in table 13.11. For *total*, the calculations are based on the entire 54 records in table 13.10, ignoring both place and season. The method is identical with that described in section 13.2 leading to lines 3 in table 13.2.

The numbers in the second line opposite *place* are derived from the place sums at the bottom of table 13.10, the calculation following the form usual in analysis of variance. For example, in line H , column N ,

$$\frac{(78.9)(40.7) + \dots + (85.1)(22.9)}{3} - \frac{(521.7)(208.4)}{18} = -47.06$$

The set of six correction terms is the same for total, place, and year. The entries opposite *season* are calculated similarly from the year sums. The remainders in the six cells of the table are the sums of squares and products for discrepancy which, in an experiment of this type, are ascribed to error. These lead to the desired regression, independent of locality and season, recorded at the bottom of the table.

The values of H and N for each place and year are substituted successively in the regression equation, leading to the estimated yields of table 13.12. One notes that the yields in 1933 are less than average, while those in 1935 are more. All the Plumpton yields are greater than average. The job is notably well done, however, when the small amount

TABLE 13.11
ANALYSIS OF VARIANCE AND COVARIANCE IN WHEAT DATA. REGRESSION
STATISTICS FOR ERROR

Source of Variation	Variate	Sums of Squares and Products		
		Height H	Number N	Yield Y
Total	H	230.53	— 0.65	341.25
Place		106.34	— 47.06	190.83
Season		6.26	26.24	8.41
Remainder (error)		117.93	20.17	142.01
Total	N		385.07	—300.75
Place			171.46	—257.03
Season			139.41	— 22.26
Remainder (error)			74.20	— 21.46
Total	Y			982.30
Place				629.22
Season				124.42
Remainder (error)				228.66

For remainder (error)

$$\begin{aligned}
 r_{HN} &= 0.2156, & r_{HY} &= 0.8648, & r_{NY} &= -0.1647 \\
 b'_{YH \cdot N} &= 0.9442, & b'_{YN \cdot H} &= -0.3683, & R^2 &= 0.8772 \\
 \hat{Y} &= 1.3147H - 0.6466N + 1.393
 \end{aligned}$$

Place + error	H	224.27	— 26.89	332.84
	N		245.66	—278.49
	Y			857.88

For place + error, $R^2 = 0.8494$

Source of Variation	Yield Sum of Squares	R^2	Errors of Estimate		
			Degrees of Freedom	Sum of Squares	Mean Square
Place + error	857.88	0.8494	13	129.20	
Error	228.66	0.8772	8	28.08	3.51
Place			5	101.12	20.22

$$F = 5.76^*, \quad d.f. = 5 \text{ and } 8$$

TABLE 13.12
ACTUAL AND ESTIMATED YIELDS OF WHEAT

Place	1933			1934			1935			Sum
	Ac- tual	Ex- pect- ed	Errors of Es- timate	Ac- tual	Ex- pect- ed	Errors of Es- timate	Ac- tual	Ex- pect- ed	Errors of Es- timate	
Seale Hayne	19.0	25.4	-6.4	32.4	30.1	2.3	26.2	26.0	0.2	
Rothamsted	22.2	26.2	-4.0	32.2	32.5	-0.3	34.7	32.3	2.4	
Newport	35.3	38.9	-3.6	43.7	43.4	0.3	40.0	37.7	2.3	
Boghall	32.8	35.3	-2.5	35.7	37.7	-2.0	29.6	26.2	3.4	
Sprowston	25.3	30.6	-5.3	28.3	29.5	-1.2	20.6	23.2	-2.6	
Plumpton	35.8	33.4	2.4	35.2	28.4	6.8	47.2	39.5	7.7	
Sums			-19.4			5.9			13.4	-0.1

of data is considered. With the accumulation of information and experience there is hope for even closer estimates.

One should ponder the fundamental character of the regression equation for error. It forms a substratum on which are built the variations in yield with place and season. It has been used heretofore for adjusting means and for testing significance. It constitutes an estimate of the regression of Y on the independent variates freed from the controlled variation in column and row sums.

In other experiments similar to this there may be occasion for tests of significance. The proper method is illustrated by testing the effect of place on yield of wheat adjusted for height and plant number. In the lower part of table 13.11 are entered the two additional steps in the test. First, the corresponding sums of squares and products in *place* and *error* having been added, the new $R^2 = 0.8494$ is calculated. The test of significance is then made in the usual manner.

13.8—Partial correlation. Related to partial regression in somewhat the same way that r is related to b in total regression (section 7.4) is the *partial correlation*. Among the variables X_1 , X_2 , and X_3 , $r_{12.3}$ is an estimate of the correlation between X_1 and X_2 in a population whose members all have the same X_3 . More realistically it is a kind of average of the correlations in groups of individuals classified on the basis of X_3 . For example, the partial correlation between age and gain *independent of initial weight* among the swine measurements of table 13.1 is approximated from table 13.4:

$$r_{Y1.2} = \frac{-9.1860}{\sqrt{(2,744.78)(0.8047)}} = -0.196$$

You see immediately that $r_{Y1.2}$ gives information like that got from $b_{Y1.2}$. In fact, one definition of partial correlation is a generalization

of that in section 7.4. If one first estimates X_1 from X_2 independent of X_3 ($b_{12 \cdot 3}$), then X_2 from X_1 ($b_{21 \cdot 3}$), he may define the partial correlation as

$$r_{12 \cdot 3} = \sqrt{b_{12 \cdot 3} b_{21 \cdot 3}},$$

the geometric mean of the two partial regressions. This emphasizes the fact that partial correlation is peculiarly appropriate for the study of relations among sets of measurements made simultaneously where there is no more reason to estimate X_1 from X_2 than to estimate X_2 from X_1 . If X_1 follows after X_2 and X_3 , or is caused by them, the partial regressions, $b_{12 \cdot 3}$ and $b_{21 \cdot 3}$, usually give the information desired.

A consequence is that one commonly wishes to know about partial correlations in a group of total correlations where regression is out of the picture. The convenient formula is then

$$r_{12 \cdot 3} = \frac{r_{12} - r_{13} r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}}$$

For an illustration we use some correlations among ear circumference e , cob circumference c , and number of rows of kernels k calculated from measurements of 900 ears of corn (2):

$$r_{ee} = 0.799, r_{ek} = 0.570, r_{ck} = 0.507$$

One might inquire if r_{ee} is conditioned by the number of kernel rows. Since

$$r_{ee \cdot k} = \frac{0.799 - (0.570)(0.507)}{\sqrt{(1 - 0.325)(1 - 0.257)}} = 0.720,$$

the conclusion would be that k has little effect on the correlation—it is much the same whether k varies or is constant. Another statement is this: the average correlation in groups of ears having the same number of kernel rows is little different from the total correlation. This statement is a generalization of the first. The information furnished by partial correlation is independent of values of the eliminated variable. This eliminated variable is not necessarily held constant.

Another way to tackle the problem of the foregoing paragraph is to compute

$$r_{ek \cdot e} = \frac{0.507 - (0.799)(0.570)}{\sqrt{(1 - 0.638)(1 - 0.325)}} = 0.105$$

Thus, cob circumference and kernel rows are practically uncorrelated within groups of ears having the same circumference.

For testing the significance of partial correlations use table 7.3 (or table 13.6, 2 variables), but enter it with degrees of freedom $n - 3$. In general, the number of degrees of freedom for a partial correlation is the number of sets of observations minus the number of variates involved.

EXAMPLE 13.11—For the pig data substitute the values of r from table 13.2 in the formula for $r_{Y1.2}$ with the result -0.190 . Compare this with the average correlation, -0.196 , within the groups of table 13.3.

EXAMPLE 13.12—Verify the fact that partial correlation is the geometric mean of two partial regressions by using results from section 13.3 and example 13.5.

EXAMPLE 13.13—In the same way verify the formula,

$$r_{12.3} = \sqrt{b'_{12.3} b'_{21.3}}$$

EXAMPLE 13.14—The data of examples 13.1 and 12.1 lend themselves to the verification of the statements made about partial correlation and of the two formulas given for $r_{12.3}$.

13.9—Four or more variates in a single group. Increasing beyond three the number of measurements on each individual though involving no new principles does introduce increasing complexity into the methods of multiple regression. Only the fundamentals will be presented here. Laboratory methods of calculation, together with more extensive discussions may be found in specialized publications (3) (4) (5) (10).

Amateurs should be cautioned not to rush headlong into regression studies involving many variables. Some people seem to think there is magic in the collection of vast amounts of data—that by some alchemy multiple regression will yield authentic information from careless measurements on heterogeneous material. The fact is that hazards increase with the extent and complexity of the investigation. Danger of the loss of data or failure of techniques, difficulty of finding homogeneous material, perplexities involved in interpretation, these are a few of the obstacles. However, do not be deterred if you have well taken measurements on carefully chosen material, and if you have definite questions whose answers lie in the methods to be used. As compared to the labor of getting the data the calculation of regression statistics is easy.

Should you have four or more variates in several subsamples of individuals, the subsamples being representative of somewhat different but related populations, methods analogous to those of sections 13.6 and 13.7 are available.

One should be increasingly cautious of including more variables in small samples. The reason is readily explained by an appeal to the geometry of regression, but we shall have to start with two variables. In a plane, two pairs of observations fix a linear regression, all additional pairs being available not only for improving the regression but for an estimate of error as well. In three-space the required number of triplets of observations to fix a regression plane is three. The analogy may be extended to four and more variables despite the limitations of our three-dimensional world. A regression in six variables would fit perfectly ($R = 1$) six sets of observations, leaving no information about error. Don't get excited, then, if you get $R = 0.9$ in a six-variable regression with only 10 sets of observations. As you will find later, that value is not even significant. One can easily be deceived by a large R if n is little

greater than the number of variates. With small samples be meticulous about all tests of significance, then be a bit skeptical besides.

13.10—Calculation of regression with four variates. For illustration we have chosen a sample of four measurements on each of 44 honey bees (8). The immediate object is to learn what dimensions, easily obtained, best indicate the length of proboscis. The latter is difficult to measure directly. The possibility of selecting bees for length of proboscis is of interest in Iowa because of the problem of pollination of red clover, the nectar in which cannot be reached by the ordinary honey bee. The variables are dry weight in mgs. (X_1), length (X_2), and width (X_3) of right forewing in mm., and length of proboscis (Y), also in mm.

The sums, sums of squares and sums of products in table 13.13 were calculated from the experimental data in the manner now familiar. The correction terms and the remaining computations of the correlation coefficients need only be mentioned. A word should be said about the number of significant figures. The correction terms are almost as large as the sums of squares of the observations, indicating small variation. In the remain-

TABLE 13.13
COMPUTATION OF CORRELATION COEFFICIENTS AMONG FOUR MEASUREMENTS
ON HONEY BEES

$N = 44$	Dry Weight X_1	Length of Wing X_2	Width of Wing X_3	Length of Proboscis Y	
Sum	576.4	422.7	144.5	290.0	
Mean	13.10	9.61	3.28	6.59	
X_1	1. ΣX^2 , etc.	7567.5240	5539.2979	1893.7740	3800.5057
	2. Correction	7550.8400	5537.3700	1892.9500	3799.0000
	3. Σx^2 , etc.	16.6840	1.9279	0.8240	1.5057
	4. $\sqrt{\Sigma x^2}$, etc.	4.0846	4.0691	1.9365	3.3759
	5. r_s		0.4738	0.4255	0.4460
X_2	1. ΣX^2 , etc.	4061.7944	1388.5203	2786.5762	
	2. Correction	4060.8020	1388.1852	2785.9773	
	3. Σx^2 , etc.	0.9924	0.3351	0.5989	
	4. $\sqrt{\Sigma x^2}$, etc.	0.9962	0.4723	0.8234	
	5. r_s		0.7095	0.7274	
X_3	1. ΣX^2 , etc.		474.7759	952.5712	
	2. Correction		474.5511	952.3864	
	3. Σx^2 , etc.		0.2248	0.1848	
	4. $\sqrt{\Sigma x^2}$, etc.		0.4741	0.3918	
	5. r			0.4716	
Y	1. ΣY^2			1912.0467	
	2. Correction			1911.3636	
	3. Σy^2			0.6831	
	4. $\sqrt{\Sigma y^2}$			0.8265	
	5. r_s			0.1260	

ders of lines 3 the last figure and in some cases the last but one have little relation to the experimental measurements. They must be retained, however, in this long series of operations in order that rounding may not impair the accuracy of the calculations.

There are now three normal equations to be solved for the three standard partial regression coefficients required in the regression equation,

$$\hat{y} = b'_{Y1.23}x'_1 + b'_{Y2.13}x'_2 + b'_{Y3.12}x'_3,$$

in which the dashed x 's represent standard measure. The symbol $b'_{Y1.23}$ is read, "the standard regression coefficient of Y on X_1 independent of X_2 and X_3 ." The normal equations are:

	X_1	X_2	X_3	Y
X_1	$b'_{Y1.23} + r_{12}b'_{Y2.13} + r_{13}b'_{Y3.12} = r_{Y1}$			
X_2	$r_{12}b'_{Y1.23} +$	$b'_{Y2.13} + r_{23}b'_{Y3.12} = r_{Y2}$		
X_3	$r_{13}b'_{Y1.23} + r_{23}b'_{Y2.13} +$	$b'_{Y3.12} = r_{Y3}$		

Notice the symmetry in these equations. The subscripts of the appropriate r 's are easily determined from those of the X 's.

Substituting the r 's from table 13.13:

	X_1	X_2	X_3	Y
X_1	$b'_{Y1.23} + 0.4738 b'_{Y2.13} + 0.4255 b'_{Y3.12} = 0.4460$ (1)			
X_2	$0.4738 b'_{Y1.23} +$	$b'_{Y2.13} + 0.7095 b'_{Y3.12} = 0.7274$ (2)		
X_3	$0.4255 b'_{Y1.23} + 0.7095 b'_{Y2.13} +$	$b'_{Y3.12} = 0.4716$ (3)		

This set of simultaneous linear equations is solved by the ordinary methods of elementary algebra. A good way to start is to divide each of the equations 1 and 2 by its coefficient of $b'_{Y3.12}$; that is, by 0.4255 and 0.7095 respectively. The results are

$$2.3502 b'_{Y1.23} + 1.1135 b'_{Y2.13} + b'_{Y3.12} = 1.0482 \quad (4)$$

$$0.6678 b'_{Y1.23} + 1.4094 b'_{Y2.13} + b'_{Y3.12} = 1.0252 \quad (5)$$

With a calculating machine the divisions are made most easily by use of the reciprocal. For example, $1/0.4255 = 2.35018$, the leading coefficient of equation 4. The products of this reciprocal by 0.4738, 0.4255, and 0.4460 are the remaining coefficients of equation 4.

Now subtract equation 3 from 4, then 5 from 4:

$$1.9247 b'_{Y1.23} + 0.4040 b'_{Y2.13} = 0.5766 \quad (6)$$

$$1.6824 b'_{Y1.23} - 0.2959 b'_{Y2.13} = 0.0230 \quad (7)$$

Multiply each of these equations by the reciprocal of its coefficient of $b'_{Y2.13}$:

$$\begin{aligned} 4.7641 b'_{Y1.23} + b'_{Y2.13} &= 1.4272 & (8) \\ - 5.6857 b'_{Y1.23} + b'_{Y2.13} &= -0.0777 & (9) \end{aligned}$$

Subtracting 9 from 8,

$$10.4498 b'_{Y1.23} = 1.5049, \quad (10)$$

whence $b'_{Y1.23} = 0.1440 \quad (11)$

This value is to be substituted back in equation 8 (or 9 or both) yielding,

$$(4.7641)(0.1440) + b'_{Y2.13} = 1.4272,$$

from which $b'_{Y2.13} = 1.4272 - 0.6861 = 0.7411$

Substituting both these b -primes in one of the original equations, 3 for example,

$$\begin{aligned} (0.4255)(0.1440) + (0.7095)(0.7412) + b'_{Y3.12} &= 0.4716 \\ b'_{Y3.12} &= 0.4716 - 0.0613 - 0.5259 = -0.1156 \end{aligned}$$

For verification, the three b -primes may be substituted in each of the remaining equations, 1 and 2. Occasionally a small r is multiplied by a small b' causing the product to have no figure in the first four decimal places. Then the equation may seem to verify the b -primes despite the fact that the small one may be in error. For that reason, it may be as well to verify them in both the original equations.

For convenience and later comparisons we bring together the following statistics:

Proboscis Length Y	Dry Weight X_1	Wing Length X_2	Wing Width X_3
Correlations of Y with X 's	0.4460.	0.7274	0.4716
Standard regressions of Y on X 's	0.1440	0.7411	-0.1155

Although not needed for this example, the regression equation is calculated following the method of section 13.3:

$$\hat{Y} = 6.59 + 0.1440 \frac{0.8265}{4.0846} (X_1 - 13.10)$$

$$+ 0.7411 \frac{0.8265}{0.9962} (X_2 - 9.61)$$

$$- 0.1155 \frac{0.8265}{0.4741} (X_3 - 3.28)$$

$$= 0.0291X_1 + 0.6148X_2 - 0.2014X_3 + 0.96$$

If the 44 estimated values, $\hat{Y}$, were computed then correlated with actual values Y , the result would be (section 13.4),

$$R^2 = (0.4460)(0.1440) + (0.7274)(0.7411) \\ + (0.4716)(-0.1155) = 0.5488$$

$$R = 0.7408$$

Since this value is little larger than $r_{Y2} = 0.7274$, small advantage has been gained by the multiple regression so far as estimating proboscis length is concerned. One might almost as well use the simple regression of proboscis length on wing length.

Another indication of the outstanding role of wing length in estimating proboscis length is the predominating size of $b'_{Y2.13} = 0.7411$. In the presence of wing length the other two variates give little information. It is notable that the total correlations do not even denote the order of importance of the three independent variables. This is a warning not to try to interpret zero order correlations except in the light of the pertinent information contained in the b -primes.

Another remark may be interpolated just here. It is usually futile to argue about cause-effect relations from information furnished by a single small correlation even though it be highly significant. Why? If this small correlation were included in a group having a large R , the b -prime corresponding to it might be found to contradict its indication, as in the example being discussed. If you wish reliable information about relations, get together a group of variables whose connections have theoretical validity then, if R is large enough to indicate fairly complete determination, examine the b -primes.

As in the first example of this chapter, it is interesting to note the change of sign from $r_{Y3} = 0.4716$ to $b'_{Y3.12} = -0.1156$. If data were sufficiently plentiful it would be possible to gather into subsamples those quadruplets of measurements in which dry weight X_1 and wing length X_2 had some constant pair of values. It would be found that the average regression of proboscis length on wing width in such subsamples would be -0.1156 . This regression would be independent of X_1 and X_2 because these two variates would be unchanging within each subsample. An interpretation may be ventured. While both wing width and wing length are affected alike by body size ($r_{23} = 0.7095$), and while length of wing and proboscis are closely related ($r_{Y2} = 0.7274$, $b'_{Y2.13} = 0.7411$), yet in groups of uniform body size and wing length the stocky bee tends to have a wide wing and short proboscis.

The significance of R is tested by referring to table 13.6, looking down the column for four variables until you get to 40 degrees of freedom (44 observations minus 4 variables), then comparing $R = 0.7408$ with the 5% and 1% values, 0.419 and 0.494. We conclude that there would be little chance of drawing our sample R from a population of uncorrelated variates. The significance of the regression coefficients will be discussed in section 13.12.

An alternative method for testing the significance of R is more enlightening though slightly less convenient. Analyze the variance of Y ($S_y^2 = 0.6831$ in table 13.13) in this manner:

Source of Variation	Degrees of Freedom	Sum of Squares		Mean Square
		Symbolical	Numerical	
Linear regression	3	$R^2 S_y^2$	0.3749	0.125
Deviations from regression	40	$(1 - R^2) S_y^2$	0.3082	0.0077
Total	43	S_y^2	0.6831	

$$F = 0.125/0.0077 = 16.2, \quad d.f. = 3 \text{ and } 40$$

The conclusions, of course, are the same as before.

13.11—Direct method of computation: 3 variates. Although correlation coefficients have been used extensively for regression calculations, they are relatively new in the history of statistics, having been introduced near the beginning of the present century. An older method involved the sums of squares and products, the normal equations for 3 variates being,

$$\begin{aligned} Sx_1^2 b_{Y1.2} + Sx_1 x_2 b_{Y2.1} &= Sx_1 y \\ Sx_1 x_2 b_{Y1.2} + Sx_2^2 b_{Y2.1} &= Sx_2 y \end{aligned}$$

These are solved directly for the b 's in the same way the correlation equations were solved for the b -primes:

$$\begin{aligned} b_{Y1.2} &= \frac{(Sx_2^2)(Sx_1 y) - (Sx_1 x_2)(Sx_2 y)}{(Sx_1^2)(Sx_2^2) - (Sx_1 x_2)^2} \\ b_{Y2.1} &= \frac{(Sx_1^2)(Sx_2 y) - (Sx_1 x_2)(Sx_1 y)}{(Sx_1^2)(Sx_2^2) - (Sx_1 x_2)^2} \end{aligned}$$

If you substitute the sums of squares and products for the pig data, you will get $b_{Y1.2} = -0.003290$ and $b_{Y2.1} = 0.007409$ as before. The regression equation becomes,

$$\hat{Y} = 1.3882 - 0.003290(X_1 - 77.050) + 0.007490(X_2 - 52.725),$$

which reduces to that calculated earlier.

The sum of squares due to regression is

$$\begin{aligned} S\hat{y}^2 &= b_{Y1.2} Sx_1 y + b_{Y2.1} Sx_2 y \\ &= (-0.003290)(6.9235) + (0.007409)(27.5408) \\ &= 0.1813, \end{aligned}$$

so that the sum of squares for deviations is

$$\begin{aligned} Sd_{Y.12}^2 &= S_y^2 - S\hat{y}^2 \\ &= 1.0228 - 0.1813 = 0.8415 \end{aligned}$$

EXAMPLE 13.15—The following data are adapted from a report by Pearl (9). The problem was to discover what factors might have been related to the intensity of the influenza epidemic of 1918. The onslaught was known to have varied widely in different cities: could this have been attributed to measurable characteristics of the cities themselves?

City	Age Distribution of Population	Latitude (Degrees)	Deaths per 100,000 From Heart Diseases	25-week Excess Mortality (per 1,000)
	X_1	X_2	X_3	Y
1. Albany	11	43	236	47
2. Atlanta	83	34	122	27
3. Baltimore	32	39	196	61
4. Boston	31	42	212	65
5. Buffalo	49	43	166	58
6. Cambridge	30	42	184	59
7. Chicago	69	42	160	38
8. Cincinnati	23	38	202	40
9. Cleveland	75	42	118	40
10. Columbus	29	40	156	32
11. Dayton	24	40	184	35
12. Fall River	73	42	164	58
13. Grand Rapids	28	43	146	15
14. Indianapolis	24	40	186	25
15. Louisville	32	38	164	36
16. Los Angeles	14	34	156	52
17. Lowell	35	43	158	51
18. Milwaukee	62	43	100	29
19. Minneapolis	55	45	114	27
20. Nashville	48	36	198	78
21. Newark	63	41	148	51
22. New Haven	34	41	186	56
23. New Orleans	52	30	224	72
24. New York	74	41	166	47
25. Oakland	17	38	188	59
26. Philadelphia	34	40	204	73
27. Pittsburgh	72	40	134	80
28. Providence	30	42	152	53
29. Rochester	27	43	204	27
30. St. Louis	45	39	142	30
31. St. Paul	69	45	118	33
32. San Francisco	34	38	238	75
33. Toledo	33	42	186	21
34. Washington	21	39	236	66
Sum	1,432	1,368	5,848	1,616

The correlations are as follows:

	X_1	X_2	Y
X_1	0.07798	-0.61118	-0.02581
X_2		-0.30774	-0.34819
X_3			0.48446

You will get $R^2 = 0.3773$ and the regression equation,

$$\hat{Y} = 0.35313X_1 - 0.95405X_2 + 0.33149X_3 + 14.026$$

Finally,

$$R^2 = Sg^2/Sy^2 \\ = 0.1813/1.0228 = 0.1773,$$

as in section 13.4, except for negligible rounding errors.

For testing the regression coefficients and for making fiducial statements there are needed 3 quantities called *multipliers*. These were introduced by Gauss (7): they are often referred to as "elements of the inverse matrix." They all have the same denominator,

$$D = (Sx_1^2)(Sx_2^2) - (Sx_1x_2)^2 \\ = (4,735.90)(5,065.98) - (3,037.55)^2 \\ = 14,765,300$$

This same denominator appeared in the formulas for the b 's; so its reciprocal, $0.677265/10^7$, may be set in the machine for multiplying the numerators of all five fractions. The multipliers are,

$$c_{11} = Sx_2^2/D = (5,065.98)(0.677265)/10^7 = 0.00034310 \\ c_{22} = Sx_1^2/D = (4,735.90)(0.677265)/10^7 = 0.00032075 \\ c_{12} = -Sx_1x_2/D = -(3,037.55)(0.677265)/10^7 = -0.00020572$$

We now calculate the variance of deviations from regression,

$$Sd_{Y.12}^2/(n - 3) = 0.8415/37 = 0.02274$$

from which, $s_{Y.12} = 0.1508$. From the above come the following standard deviations and values of t :

$$\text{For } b_{Y1.2}: s = s_{Y.12}\sqrt{c_{11}} = 0.002793, t = 0.003290/0.002793 = 1.18 \\ \text{For } b_{Y2.1}: s = s_{Y.12}\sqrt{c_{22}} = 0.002701, t = 0.007409/0.002701 = 2.74$$

The values of t are identical with those for the corresponding b -primes: it is a matter of preference which method is used for making the test.

Since we now have the standard deviation of $b_{Y1.2}$, we may set 95% fiducial limits,

$$-0.003290 \pm (2.026)(0.002793);$$

that is, from -0.008949 to 0.002369 . This shows the unreliability of our estimate of this partial regression—it might be positive in the population instead of negative as in the sample. Of the other coefficient the limits are from 0.001937 to 0.012881 pound per day per pound initial weight.

For the fiducial limits of any $\hat{Y}$, the variance is computed by,

$$s_{Y.12}^2(1/n + c_{11}x_1^2 + c_{22}x_2^2 + 2c_{12}x_1x_2)$$

where x_1 and x_2 are deviations from their respective means. As an example, let us calculate the limits for the rate of gain estimated for pigs weighing 60 pounds at 60 days of age. Substituting in the regression equation, the predicted rate is

$$\hat{Y} = 1.3882 - 0.003290(60 - 77.050) + 0.007409(60 - 52.725) \\ = 1.50 \text{ pound per day,}$$

with the variance,

$$0.02274[1/40 + (0.00034310)(-17.050)^2 + (0.00032075)(7.275)^2 \\ + 2(-0.00020572)(-17.050)(7.275)] = 0.004384$$

Thus, the standard error is 0.0662. Finally, the 95% limits are $1.50 \pm (2.026)(0.0662)$; that is, from 1.37 to 1.63 pounds per day.

The variance of the estimated rate of an individual pig could be found, as in section 6.9, by adding $s_{Y \cdot 12}^2 = 0.02274$ to 0.004384.

13.12—The Gauss multipliers with more than 3 variates. As the number of variates increases the computations grow heavier, but no new features are added. For 4 variates, the normal equations in the direct method are,

$$\begin{aligned} Sx_1^2b_{Y1.23} + Sx_1x_2b_{Y2.13} + Sx_1x_3b_{Y3.12} &= Sx_1y \\ Sx_1x_2b_{Y1.23} + Sx_2^2b_{Y2.13} + Sx_2x_3b_{Y3.12} &= Sx_2y \\ Sx_1x_3b_{Y1.23} + Sx_2x_3b_{Y2.13} + Sx_3^2b_{Y3.12} &= Sx_3y \end{aligned}$$

After substitution of sample data for the sums of squares and products, the three b 's may be solved for by the usual methods; the computations of the regression equation, $S\hat{y}^2$, $Sd_{Y \cdot 123}^2$, and R^2 parallel those for 3 variates. Instead of giving these details, we shall show how to compute the Gauss multipliers and to use them for completing the regression calculations. This may be done in either the correlation or the direct method: we shall do the latter, then show later some of the relations.

The equations giving the multipliers are got from the normal equations by replacing the b 's by c 's and the right members successively by 1, 0, 0; 0, 1, 0; and 0, 0, 1; then solving each set for the c 's. For example, in the first set, using the bee data in table 13.13:

$$\begin{aligned} 16.6840 c_{11} + 1.9279 c_{21} + 0.8240 c_{31} &= 1 \\ 1.9279 c_{11} + 0.9924 c_{21} + 0.3351 c_{31} &= 0 \\ 0.8240 c_{11} + 0.3351 c_{21} + 0.2248 c_{31} &= 0 \end{aligned}$$

Solving as before, $c_{11} = 0.07892$, $c_{21} = -0.11202$, and $c_{31} = -0.12230$.

Solving the second set, with 0, 1, 0 on the right, $c_{12} = -0.11202 = c_{21}$, $c_{22} = 2.18790$, $c_{32} = -2.85079$. The third set with 0, 0, 1 on the right, gives $c_{13} = -0.12230 = c_{31}$, $c_{23} = -2.85079 = c_{32}$, and $c_{33} = 9.14624$. These elements are entered in table 13.14 along with the sums of products of Y by the X 's. At first, the computation may seem forbidding, but actually it is not difficult. Some short-cuts will be shown below after the utility of the multipliers has been presented. The following statistics may now be calculated:

1. *The regression coefficients.* One of them is taken from each column of multipliers in the table. As an example, from solution 2,

$$\begin{aligned} b_{Y2.13} &= c_{12}Sx_1y + c_{22}Sx_2y + c_{32}Sx_3y \\ &= (-0.11202)(1.5057) + (2.18790)(0.5989) \\ &\quad + (-2.85079)(0.1848) = 0.61484 \end{aligned}$$

TABLE 13.14

THREE SETS OF MULTIPLIERS TOGETHER WITH SUMS OF y -PRODUCTS AND RESULTING REGRESSION COEFFICIENTS. SEE DATA

Multiplier	Solution Number			Sums of y -products
	1	2	3	
c_1	$c_{11} = 0.07892$	$c_{12} = -0.11202$	$c_{13} = -0.12230$	1.5057
c_2	$c_{21} = -0.11202$	$c_{22} = 2.18790$	$c_{23} = -2.85079$	0.5989
c_3	$c_{31} = -0.12230$	$c_{32} = -2.85079$	$c_{33} = 9.14624$	0.1848
Regression Coefficients	0.02914	0.61484	-0.20126	

The regression equation is then computed as usual.

2. *The sum of squares attributable to regression.*

$$\begin{aligned} S\hat{y}^2 &= b_{Y1.23}Sx_1y + b_{Y2.13}Sx_2y + b_{Y3.12}Sx_3y \\ &= (0.02914)(1.5057) + (0.61484)(0.5989) + (-0.20126)(0.1848) \\ &= 0.3749 \end{aligned}$$

3. *The sum of the squares of deviations from regression.*

$$Sd_{Y.123}^2 = S_y^2 - S\hat{y}^2 = 0.6831 - 0.3749 = 0.3082$$

4. *The variance and standard error of estimate.*

$$s_{Y.123}^2 = Sd_{Y.123}/d.f. = 0.3082/37 = 0.007705. \quad s_{Y.123} = 0.08778$$

5. *The multiple correlation coefficient.*

$$R^2 = S\hat{y}^2/S_y^2 = 0.3749/0.6831 = 0.5488. \quad R = 0.7408$$

6. *The standard deviations of the regression coefficients.* As an example, for $b_{Y2.13}$:

$$s = s_{Y.123}/c_{22} = (0.08778)(1.4792) = 0.01298$$

The testing of significance and setting of fiducial limits now proceeds as usual.

7. *The standard deviation of the difference between two regression coefficients.* This is seldom wanted, but if there were interest in testing the hypothesis that the regression on wing length and on wing width are the same, the variance of the difference is

$$\begin{aligned} s_{Y.123}^2(c_{11} + c_{22} - 2c_{12}) \\ &= 0.007705(0.07892 + 2.18790 - 2[-0.11202]) \\ &= 0.019192 \end{aligned}$$

Hence, the standard error is 0.1385, so that the difference, $0.61484 - (-0.20126) = 0.81610$, is tested by $t = 0.81610/0.1385 = 5.89$, $d.f. = 37$.

8. *The standard error of $\hat{Y}$.* The variance of this estimate is

$$s_{Y \cdot 123}^2(1/n + c_{11}x_1^2 + c_{22}x_2^2 + c_{33}x_3^2 + 2c_{12}x_1x_2 + 2c_{13}x_1x_3 + 2c_{23}x_2x_3),$$

where the x 's are deviations from their respective means. The standard error may be used for calculating fiducial limits of $\hat{Y}$ and (after the usual modification) of Y .

EXAMPLE 13.16—In the direct method, it is advantageous to code the data so as to make the ranges of the independent variates as nearly alike as feasible. For the influenza data of example 13.15, this may be done as follows: (i) subtract 30° from each X_2 , then multiply by 6; (ii) divide each X_3 by 2. Since this involves no rounding, the final results will be unaltered. The values of the variates for the first two cities are:

Albany: 11, 78, 118, 47
Atlanta: 83, 24, 61, 27

Complete the table.

EXAMPLE 13.17—Using the coded data of the preceding example, compute the sums of squares and products:

	X_1	X_2	X_3	Y
X_1	13,657.53	1,008.35	-7,501.00	- 310.12
X_2		12,244.24	-3,576.00	-3,961.41
X_3			11,028.00	5,231.00
Y				10,572.47

EXAMPLE 13.18—From the foregoing sums of squares and products, calculate the regression coefficients,

$$0.35316, \quad -0.15898, \quad 0.66300$$

To compare these with the coefficients in example 13.15, you will have to reverse the coding process, dividing -0.95405 by 6 and multiplying 0.33149 by 2. The discrepancies in the fifth place of decimals are the accumulated effects of rounding in the two solutions of the normal equations.

13.13—Deletion of a variate. After a regression is completed, the value of a variate may be questioned and its omission proposed. The computations with the reduced number of variates may, of course, be carried out anew, but desired information is often to be had more easily by means of the formulas below (6).

For the bees, the regression on width of wing is small and nonsignificant. If this variate, X_3 , is deleted, the sum of squares due to regression is decreased by

$$b_{Y3 \cdot 12}^2/c_{33} = (-0.20126)^2/9.14624 = 0.0044,$$

leaving for the reduced regression,

$$S\hat{y}_{12}^2 = 0.3749 - 0.0044 = 0.3705,$$

the subscripts indicating that $\hat{y}$ is now estimated from X_1 and X_2 only. The corresponding multiple correlation coefficient is

$$R_{Y \cdot 12}^2 = S\hat{y}_{12}^2/Sy^2 = 0.3705/0.6831 = 0.5424, \\ R_{Y \cdot 12} = 0.7365, \text{ d.f.} = 38$$

TABLE 13.15
COMPUTATION OF THE MULTIPLIERS FOR THE BEE DATA OF TABLE 13.13. ENTRIES ARE ROUNDED TO 4 DECIMALS, BUT THE ORIGINAL CALCULATION REQUIRED 6

Line	Directions	Coefficients of			Solution Number		
		c_1	c_2	c_3	1	2	3
(1)	Copy from lines number 3, table 13.13	16.6840	1.9279	0.8240	1	0	0
(2)		1.9279	0.9924	0.3351	0	1	0
(3)		0.8240	0.3351	0.2248	0	0	1
(4)	Multiply (1) by 1/0.8240	20.2475	2.3397	1.0000	1.2136	0.	0.
(5)	Multiply (2) by 1/0.3351	5.7532	2.9615	1.0000	0.	2.9842	0.
(6)	Multiply (3) by 1/0.2248	3.6655	1.4901	1.0000	0.	0.	4.4484
(7)	(4) - (5)	14.4943	-0.6218		1.2136	-2.9842	0.
(8)	(5) - (6)	2.0877	1.4708		0.	2.9842	-4.4484
(9)	Multiply (7) by 1/(0.6218)	23.3095	-1.0000		1.9517	-4.7991	0.
(10)	Multiply (8) by 1/1.4708	1.4194	1.0000		0.	2.0289	-3.0244
(11)	(9) + (10)	24.7289			1.9517	-2.7702	-3.0244
(12)	Multiply (11) by 1/24.7289				0.0789	-0.1120	-0.1223
(13)	Substitute c_1 in (9)	c_1		c_2	-0.1120	2.1879	-2.8508
(14)	Substitute c_1 and c_2 in (5) and (4)			c_3	-0.1223	-2.8508	9.1462

If the reduced regression is decided upon, the two partial regression coefficients are calculated easily:

$$\begin{aligned}b_{Y1.2} &= b_{Y1.23} - c_{13}(b_{Y3.12}/c_{33}) = 0.02645 \\b_{Y2.1} &= b_{Y2.13} - c_{23}(b_{Y3.12}/c_{33}) = 0.55212\end{aligned}$$

Or, one may prefer to calculate the multipliers in the reduced regression. These are,

$$\begin{aligned}(c_{11})_R &= c_{11} - c_{13}^2/c_{33} = 0.07728 \\(c_{22})_R &= c_{22} - c_{23}^2/c_{33} = 1.29934 \\(c_{12})_R &= c_{12} - c_{13}c_{23}/c_{33} = -0.15014\end{aligned}$$

From these, all the statistics in the 3-variable regression may be calculated, including those given above.

Formulas for deleting a different variate may be inferred from those given: only keep in mind the number of the variate being eliminated and that of the one for which the statistic is being computed. Analogous formulas are easily written for deletion in a regression having more than 4 variates.

13.14—Calculation of the multipliers. Dozens of methods are available, each offering some advantages. I shall do no more than indicate how the elementary solution of simultaneous equations may be systematized. For the occasional user, with a moderate number of variates, no time is saved by the specialized methods because the procedures must be re-memorized each time. Those who do many regressions will no doubt adopt one of the more abbreviated schemes (3) (4) (5) (10) already mentioned.

In table 13.15, the solutions of the three sets of equations are carried on simultaneously with this advantage: each divisor is applied to an entire line in the table by setting its reciprocal in the machine and multiplying successively by all the necessary coefficients. Thus, in line 4, the reciprocal of the divisor, 0.8240, is entered at the proper place in the table ($1/0.8240 = 1.2136$), and is then put on the keyboard for multiplication by 16.6840 and 1.9279.

Those familiar with simultaneous equations will have no difficulty in following the directions. If you wish to verify the fourth place of decimals you will have to carry all calculations to six places. This is necessary in order to make $c_{12} = c_{21}$, etc. Otherwise, rounding will prevent exact verification.

After one becomes familiar with the process he will agree, I think, that the computation is not excessive if judged by the amount of information obtained.

EXAMPLE 13.19—Using the sum of squares and products in example 13.17, calculate the multipliers, c :

0.00011944	0.00001534	0.00008621
	0.00009219	0.00004033
		0.00016240

13.15—Relations between the correlation and direct methods.

Each method leads to a complete set of regression constants: the user, then has entire freedom of choice.

Attention has been called to the fact that either set of partial regression coefficients can be calculated from the other. For example,

$$b_{Y1.23} = b'_{Y1.23} \frac{\sqrt{S_y^2}}{\sqrt{S_{x_1}^2}}$$

The quantities, R and $s_{Y.123}$, are derived as parts of either procedure.

The multipliers, c' , in the correlation method may be computed from the normal equations in section 13.10 by changing the b -primes to c -primes and replacing the right members by 1, 0, 0, etc. The following relations will enable one to transfer from one set to another:

$$\begin{aligned} c_{11} &= c'_{11}/S_{x_1}^2, \text{ etc.} \\ c_{12} &= c'_{12}/S_{x_1x_2}, \text{ etc.} \end{aligned}$$

The c -primes may be entered in a table like 13.14, with the correlations, r_{Y1} , r_{Y2} , etc., in the column at the right; then the columns may be combined as before, the sums of products being the b -primes.

For testing significance of the b -primes, the quantity,

$$v = \frac{1 - R^2}{n - m} = \frac{1 - 0.5488}{44 - 4} = 0.01128,$$

replaces $s_{Y.123}$ so that the variance of $b'_{Y1.23}$ is vc'_{11} . The ensuing values of t are identical; that is

$$\frac{b'_{Y1.23}}{\sqrt{vc'_{11}}} = \frac{b_{Y1.23}}{s_{Y.123}\sqrt{c_{11}}}$$

With one exception, the formulas for deleting a variate apply to primed b 's and c 's as well as to the un-primed. The exception: if X_3 is deleted, R^2 (instead of $R^2S_y^2$) is decreased by $b'_{Y3.12}^2/c'_{33}$.

Mastery of the computational methods enables the investigator to get desired information without excess calculation. If it is the regression equation with an overall test of significance that is needed, the direct solution of the normal equations is indicated. A detailed study of correlation and regression coefficients calls for the correlation method, either with or without the multipliers. On the other hand, the direct method with multipliers leads to tests of significance and fiducial statements involving $\bar{Y}$, Y , and the b 's. For maximum flexibility in examining the regression, start by calculating the multipliers.

EXAMPLE 13.20—From the correlations in example 13.15, compute the multipliers, c' :

1.63117	0.19850	1.05803
	1.12881	0.46876
		1.79093

EXAMPLE 13.21—If you solve the normal equations in section 13.3, substituting c' for b' together with 1,0 and 0,1 in the right members, you will get the multipliers,

$$c'_{11} = \frac{1}{1 - r_{12}^2} = c'_{22}$$

$$c'_{12} = - \frac{r_{12}}{1 - r_{12}^2}$$

The formulas in section 13.5 may be derived from these multipliers.

EXAMPLE 13.22—Delete X_2 from the influenza regression. All the results may be verified by calculating the statistics of the reduced regression, using the formulas which have been given for the multipliers (section 13.11 and example 13.21).

REFERENCES

1. M. M. BARNARD. *Journal of Agricultural Science*, 26:456 (1936).
2. ARTHUR M. BRUNSON AND J. G. WILLIER. *Journal of the American Society of AGRONOMY*, 21:912 (1929).
3. M. H. DOOLITTLE. *United States Coast and Geodetic Survey Report*, 1878, 115.
4. PAUL S. DWYER. *Journal of the American Statistical Association*, 37:441 (1942). Contains bibliography.
5. MORDECAI EZEKIEL. *Methods of Correlation Analysis*. John Wiley and Sons, Inc., New York (1930).
6. R. A. FISHER. *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh (1925).
7. CARL FRIEDRICH GAUSS. *Theoria Combinationis Observatorum, Pars Posterior* (1821), *Werke*, 4:31; Supplement (1826), *Werke*, 4:71.
8. ROY A. GROUT. *Iowa Agricultural Experiment Station Research Bulletin* 218 (1937).
9. RAYMOND PEARL. *Public Health Reports* 34:1743 (1919); 36:289 (1921).
10. H. A. WALLACE AND GEORGE W. SNEDECOR. *Correlation and Machine Calculation*. Revised edition. Iowa State College Official Publication 30, No. 4 (1931). This publication, now out of print, will be replaced as soon as practicable.
11. F. YATES. *Journal of the Ministry of Agriculture*, 43:156 (1936).

Chapter 14

CURVILINEAR REGRESSION

14.1—Introduction. Although linear regression is adequate for many needs, it is a matter of common observation that some variates are not connected by so simple a relation. The discovery of a precise description of the concomitant variation of two or more quantities is one of the problems of *curve fitting*, known as *curvilinear regression*. From this general view the fitting of the straight line is a special case, the simplest and indeed the most useful.

The motives for fitting curves to nonlinear data are various. Sometimes there is the desire to make a good estimate of the dependent variate for any particular value of the independent. This may involve the smoothing of irregular data and the interpolation of estimated Y 's for values of X not contained in the observed series. Sometimes the objective is to test or discover a law relating the variates, such as a growth curve. At yet other times, the form of the relationship is of little interest; the end in view is merely the elimination of inaccuracies which nonlinearity of regression may introduce into a correlation coefficient or an experimental error.

The problem of fitting curves to data has been approached from many directions. Except when a law is being tested, the chief difficulty is the selection of a suitable equation. Graphical devices of rectification (1) and more theoretical methods (3) (7) (14) are useful and readily available to one who has some acquaintance with mathematics. In the main, we shall confine our attention to the exponential growth curve and to the polynomial.

14.2—The exponential growth curve. A characteristic of some of the simpler growth phenomena is that the increase at any moment is proportional to the size already attained. This is sometimes called the law of compound interest. During one phase in the growth of a culture of bacteria, the numbers of organisms follow such a law. The relation is nicely illustrated by the dry weights of chick embryos at ages 6 to 16 days (12) recorded in table 14.1. The graph of the weights in figure 14.1 ascends with greater rapidity as age increases, the regression equation being of the form commonly associated with such growth,

$$W = (A)(B^X),$$

where A and B are constants to be evaluated. Applying logarithms to the equation,

$$\begin{aligned}\log W &= \log A + (\log B)X \\ \text{or } Y &= a + bX,\end{aligned}$$

TABLE 14.1
 DRY WEIGHTS OF CHICK EMBRYOS FROM AGES 6 TO 16 DAYS,
 TOGETHER WITH COMMON LOGARITHMS

Ages in Days X	Dry Weight, W (grams)	Common Logarithm of Weight Y
6	0.029	-1.538*
7	0.052	-1.284
8	0.079	-1.102
9	0.125	-0.903
10	0.181	-0.742
11	0.261	-0.583
12	0.425	-0.372
13	0.738	-0.132
14	1.130	0.053
15	1.882	0.275
16	2.812	0.449

* From the table of logarithms, one reads $\log 0.029 = 8.462 - 10$ or $\bar{2}.462$, which means $-2 + 0.462 = -1.538$.

where $Y = \log W$, $a = \log A$, and $b = \log B$. This means that if $\log W$ instead of W is plotted against X , the graph will be linear. By the device of using the logarithm instead of the quantity itself, the data are said to be *rectified*.

The values of $Y = \log W$ are set out in the last column of the table and are plotted opposite X in the figure. The regression equation, computed in the familiar manner from the columns X and Y in the table, is

$$Y^* = 0.1959X - 2.689$$

The regression line fits the data points with unusual fidelity, the correlation between Y and X being 0.9992. The conclusion is that the chick embryos, as measured by dry weight, are growing in accord with the exponential law, the logarithm of the dry weight increasing at the estimated uniform rate of 0.1959 per day.

Often there is little interest in the regression equation itself, the objective being merely to learn if the data follow the exponential law. The graph, then, may furnish sufficient information. If so, the use of *semi-logarithmic* or *ratio* graph paper obviates the necessity for looking up the logarithms of Y . The horizontal rulings on the paper are drawn to such a scale that the plotting of the original data results in a straight line if the data follow the exponential growth law. Ratio paper may be purchased at most stationery shops.

It is sometimes considered desirable to write the exponential growth law in the form,

$$W = Ae^{bx},$$

* The distinction between Y and estimated Y is usually obvious. $\hat{Y}$ will be used only when the error of estimate, $Y - \hat{Y}$, is to be emphasized.

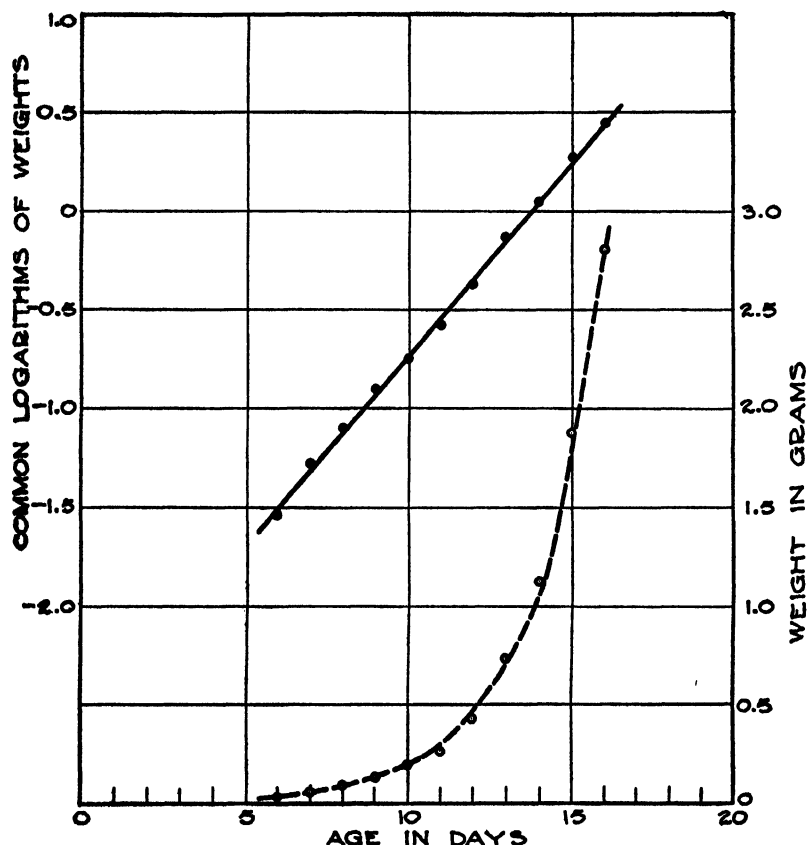


FIG. 14.1—Dry weights of chick embryos at ages 6–16 days with fitted curves.

$$\begin{aligned}\text{Uniform scale: } W &= 0.002046(1.57)^X \\ \text{Logarithmic scale: } Y &= 0.1959X - 2.689\end{aligned}$$

where e is the base of the natural or Napierian system of logarithms, an irrational number whose approximate value is 2.718. If this form is decided upon, the convenient way is to use natural logarithms; then,

$$\log_e W = \log_e A + bX$$

It may be shown that the regression coefficient, b , in this equation is the *relative* rate of increase of W ; that is, the rate of increase per unit X per unit W . This relative rate is 2.3026 ($= \log_{10} 10$) times the rate indicated by the slope of the regression line in common logarithms. Thus, for the chick embryos, the relative rate is $(2.3026)(0.1959) = 0.4511$ gram per day per gram. It is clear that the information about relative rate of growth can be deduced from either form of the equation. One may just

as well use common logarithms if he finds them more convenient, then get the correct relative rate by multiplication as above. There are no other advantages in the natural logarithms unless tables of natural logarithms are more readily available than those of common logarithms.

If there is a necessity for converting the equation, $\log W = 0.1959X - 2.689$, into the exponential form,*

$$W = (0.002046)(1.57)^X,$$

or the form†

$$W = (0.002046)e^{0.4511X},$$

it must be remembered that the sum of squares of the errors of estimate is not a minimum for the exponential. It is the sum of the squares of $\log W - \text{estimated } \log W$ which was minimized. If it is important to reduce to a minimum the sum of squares of deviations from the exponential, methods are available (14, VI, and 3, as examples). Ordinarily there is no object in reconverting the equation to the exponential form, or in making the slight improvement in the constants even if it is so converted.

Workers in other lines may wish to test different theoretical regressions. Occasionally the simple inverse proportion is indicated,

$$Y = 1/X$$

Other possibilities are $Y = \log X$, $Y = \sqrt{X}$ and $\log Y = a + b \log X$. The testing of the applicability of the proposed law should first be done graphically. Should the data appear to be rectified by the transformation of one or both the variates, then proceed with the regression computations if desired. For testing the last of the above equations, logarithmic paper is available, both vertical and horizontal rulings being in the logarithmic scale. For directions to the use of graph paper with special scales, see reference (8).

EXAMPLE 14.1—J. W. Gowen and W. C. Price counted the number of lesions of *Aucuba mosaic virus* developing after exposure to X-rays for various times. The following record is from as yet unpublished data made available through courtesy of the investigators.

Minutes exposure	0	3	7.5	15	30	45	60
Count in hundreds	271	226	209	108	59	29	12

Plot the count as ordinate, then plot its logarithm. Derive the regression, $Y = 2.432 - 0.02227X$, where Y is the logarithm of the count and X is minutes exposure.

* To get this form, note that $0.1959 = \log 1.57$ and $2.689 = \log 488.65$, hence $\log W = (\log 1.57)(X) - \log 488.65 = \log 1.57^X - \log 488.65$

$$= \log \frac{1.57^X}{488.65} = \log (0.002046)(1.57^X).$$

† This comes from the first form because $\log_e 1.57 = 0.4511$, whence $1.57 = e^{0.4511}$.

EXAMPLE 14.2—Repeat the fitting of the last example using natural logarithms. Verify the fact that the rate of decrease of hundreds of lesions per minute per hundred is $(2.3026)(0.02227) = 0.05128$.

EXAMPLE 14.3—If the meaning of relative rate isn't quite clear, try this approximate method of computing it. The increase in weight of the chick embryo during the thirteenth day is $1.130 - 0.738 = 0.392$ gram; that is, the average rate during this period is 0.392 gram per day. But the average weight during the same period is $(1.130 + 0.738)/2 = 0.934$ gram. The relative rate, or rate of increase of each gram, is therefore $0.392/0.934 = 0.42$ gram per day per gram. This differs from the average obtained in the whole period from 6 to 16 days, 0.4511, partly because the average weight as well as the increase in weight in the thirteenth day suffered some sampling variation, and partly because the correct relative rate is based on weight and increase in weight at any instant of time, not on day averages.

EXAMPLE 14.4—Geddes (6) gives single figure estimates of the baking quality of straight grade flour (Q) after being heated at 170°F . for various numbers of hours (T). Pairs of values (T, Q) are as follows: 0.25, 93; 0.50, 71; 0.75, 63; 1.0, 54; 1.5, 43; 2.0, 38; 3.0, 29; 4.0, 26; 6.0, 22; 8.0, 20. Plot Q on the vertical axis against T on the horizontal. Setting $\log T = X$ and $\log Q = Y$, plot Y against X . Derive the regression, $Y = 1.7116 - 0.4678X$.

EXAMPLE 14.5—Placing $Y = 1/Q$ in the last example, plot Y against T . Since the graph is not linear, try $Y = \frac{1}{Q - 14}$. Compute the regression of the rectified data,

$$Y = 0.01996T + 0.00546, \text{ from which } Q = \frac{1}{0.01996T + 0.00546} + 14. \text{ Running (14,}$$

VI) gives a graphical method for estimating the number that must be subtracted from W in order to get a straight regression. Can you improve on 14? How many degrees of freedom remain for testing with the mean square of errors of estimate?

EXAMPLE 14.6—Decker and Andre (2) investigated the mortality of chinch bugs exposed at a temperature of -12.2°C . for various periods:

Hours exposure	0.25	0.5	1	4	12	24	48	72
Percentage mortality	59	63	65	68	70	73	74	75

Let Y = percentage mortality and X = logarithm of hours exposure. Fit the equation, $Y = 6.051X + 64.1$. To plot these data on logarithmic paper, put the logarithmic scale horizontal and plot hours on it.

EXAMPLE 14.7—In an experiment on wheat in Australia, fertilizers were applied at various levels with these resulting yields:

Level	X	0	10	20	30	40	60
Yield	y	26.2	34.0	36.3	37.8	38.6	38.9

Fit to these data the Mitscherlich equation (16),

$$Y = a - bX,$$

where $Y = \log(39 - y)$. The yield, 39, is taken as the maximum yield irrespective of the amount of fertilizer. The fitted equation is $Y = 0.01434X - 0.2294$.

14.3—The second degree polynomial. Faced by nonlinear regression, one often has no knowledge of a theoretical equation to use. In many instances the second degree polynomial,

$$Y = a + bX + cX^2,$$

will be found to fit the data satisfactorily. The graph is a parabola whose axis is vertical, but usually only small segments of such a parabola appear in the process of fitting. Instead of rectifying the data a third variate is added, the square of X . This introduces the methods of multiple regression. The calculations proceed exactly as in chapter 13, X and X^2 being the two independent variates. It need only be remarked that $\sqrt{X}$, $\log X$, or $1/X$ might have been added instead of X^2 if the data had required it.

To illustrate the method and some of its applications, we present the data on wheat yield and protein content (10) in table 14.2 and figure 14.2. The investigator wished to estimate the protein content for various yields. We shall also test the significance of the departure from linearity.

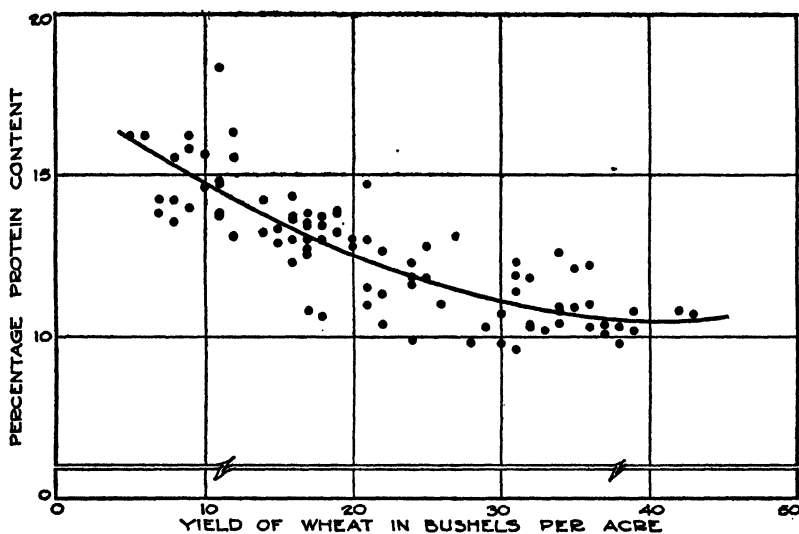


FIG. 14.2—Regression of protein content on yield in wheat, 91 plots.
 $Y = 17.703 - 0.3415X + 0.004075X^2$

The second column contains the squares of the yields in column 1. The squares are treated in all respects like a third variate in multiple regression. The computations described in chapter 13 yield the regression statistics of table 14.3. The regression equation, calculated as usual,

$$Y = 17.703 - 0.3415X + 0.004075X^2,$$

is plotted in the figure. At small values of yield the second degree term with its small coefficient is scarcely noticeable, the graph falling away

TABLE 14.2
PERCENTAGE PROTEIN CONTENT (Y) AND YIELD (X) OF WHEAT
FROM 91 PLOTS*

Yield, Bushel Per Acre X	Square X^2	Percentage Protein Y	Yield, Bushel Per Acre X	Square X^2	Percentage Protein Y
43	1,849	10.7	19	361	13.9
42	1,764	10.8	19	361	13.2
39	1,521	10.8	19	361	13.8
39	1,521	10.2	18	324	10.6
38	1,444	10.3	18	324	13.0
38	1,444	9.8	18	324	13.4
37	1,369	10.1	18	324	13.7
37	1,369	10.4	18	324	13.0
36	1,296	10.3	17	289	13.4
36	1,296	11.0	17	289	13.5
36	1,296	12.2	17	289	10.8
35	1,225	10.9	17	289	12.5
35	1,225	12.1	17	289	12.7
34	1,156	10.4	17	289	13.0
34	1,156	10.8	17	289	13.8
34	1,156	10.9	16	256	14.3
34	1,156	12.6	16	256	13.6
33	1,089	10.2	16	256	12.3
32	1,024	11.8	16	256	13.0
32	1,024	10.3	16	256	13.7
32	1,024	10.4	15	225	13.3
31	961	12.3	15	225	12.9
31	961	9.6	14	196	14.2
31	961	11.9	14	196	13.2
31	961	11.4	12	144	15.5
30	900	9.8	12	144	13.1
30	900	10.7	12	144	16.3
29	841	10.3	11	121	13.7
28	784	9.8	11	121	18.3
27	729	13.1	11	121	14.7
26	676	11.0	11	121	13.8
26	676	11.0	11	121	14.8
25	625	12.8	10	100	15.6
25	625	11.8	10	100	14.6
24	576	9.9	9	81	14.0
24	576	11.6	9	81	16.2
24	576	11.8	9	81	15.8
24	576	12.3	8	64	15.5
22	484	11.3	8	64	14.2
22	484	10.4	8	64	13.5
22	484	12.6	7	49	13.8
21	441	13.0	7	49	14.2
21	441	14.7	6	36	16.2
21	441	11.5	5	25	16.2
21	441	11.0			
20	400	12.8			
20	400	13.0			

* Read from published graph. This accounts for the slight discrepancy between the correlation we got and that reported by the author.

TABLE 14.3
REGRESSION STATISTICS FOR PROTEIN CONTENT OF WHEAT

Variate	Zero Order Correlations	
	X^2	Y
X	0.9804	-0.8063
X^2		-0.7506

$$\begin{aligned} S_y^2 &= 315.75 \\ R^2 &= 0.69111 \\ R &= 0.8313 \end{aligned}$$

$$\begin{aligned} \text{Remainder after linear regression} &= (1 - r_{xy}^2)(S_y^2) = 110.48 \\ \text{Remainder after curved regression} &= (1 - R^2)(S_y^2) = 97.53 \end{aligned}$$

almost like a straight line. Toward the right, however, the term in X^2 has bent the curve to practically a horizontal direction.

It is interesting to observe the effect of curvilinearity of regression on the correlations. If linear regression is assumed, the correlation between yield and protein content is -0.8063 ; but when the second degree term is introduced, the correlation between Y and $\hat{Y}$ becomes $R = 0.8313$. Is this small increase in absolute value significant of a curvilinear population regression? We set up the hypothesis of linear population regression, calculate the sum of squares of errors of estimate at the bottom of table 14.3, then enter the result along with its degrees of freedom ($91 - 2 = 89$) in table 14.4. In like manner, the sum of squares and degrees of freedom

TABLE 14.4
TEST OF SIGNIFICANCE OF DEPARTURE FROM LINEAR REGRESSION

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Deviations from linear regression	89	110.48	
Deviations from curved regression	88	97.53	1.11
Curvilinearity of regression	1	12.95	12.95**

$$F = 12.95/1.11 = 11.7$$

for curvilinear regression are calculated and recorded. The reduction in sum of squares, tested against the mean square remaining after curvilinear regression, proves to be significant. The hypothesis of linear regression is abandoned, therefore: there is said to be a significant curvilinearity in the regression.

For any given yield, say 20 bushels per acre, the expected protein content is

$$Y = 17.703 - (0.3415)(20) + (0.004075)(20)^2 = 12.51\%$$

In this way, the regression equation is useful for estimating and interpolating. Fiducial statements and tests of hypotheses are made as in chapter 13.

It may be thought that there is a great chasm between the rectification of data in the last section and the treatment of curvilinear regression in this. The distinction, however, is really superficial. If a three-dimensional graph were made, X and X^2 being plotted on separate axes, the regression surface would be a plane. In that sense the data would be rectified as before. It is only the limitations of the two-dimensional graph that make the methods seem different.

As always in regression, either linear or curved, one should be wary of extrapolation. The data may be incompetent to furnish evidence of trend beyond their own range. Looking at figure 14.1, one might be tempted by the excellent fit to assume the same growth rate before the sixth day and after the sixteenth. The fact is, however, that there were rather sharp breaks in the rate of growth at both these days. To be useful, extrapolation requires extensive knowledge and keen thinking.

EXAMPLE 14.8—Fit a second degree polynomial to the data of example 14.1. The result is $Y = 262.245 - 9.851X + 0.09681X^2$, not a very good fit.

EXAMPLE 14.9—Decker and Andre (2) exposed eight lots of 500 chinch bugs at -6.6°C . for varying intervals of time. The resulting mortality was as follows:

Days exposure	1	3	3.5	8	13.5	16	21	28
Percentage mortality	0.8	3.6	5.8	11.6	22.8	44.6	67.6	92.0

Fit a second degree polynomial to learn if it reveals a significant deviation from linear regression. $F = 6.09$, $d.f. = 1$ and 5 , shows that the departure from linear regression, as tested by the fitted parabola, is just short of significance.

EXAMPLE 14.10—You may think that the parabola is not a suitable equation to fit to the data of the last example. Try putting Y equal to the logarithms of the mortality percentages, then fitting $Y = a + bX$, where X is days of exposure.

How will you compare the fit of this curve with that of the parabola? For the latter, the mean square deviation from second degree regression was 32.2 with $d.f. = 5$. To get a comparable figure from the logarithmic fitting, transform each Y to its anti-logarithm P and calculate the deviations from regression. The mean of their squares is 1,270, far greater than the corresponding figure for deviation from second degree regression.

14.4—Test of deviation from linear regression. In the foregoing section this test was made in table 14.4 after the curvilinear regression had been computed. If the data occur in groups, and if it is only deviation from linearity of regression that is interesting, the test can be made rather simply. The method is illustrated by use of the data in table 14.5, made available through the courtesy of B. J. Vos and W. T. Dawson. Four rates of injection were used, each double the preceding. The particular question under investigation was whether the regression of Y on X is curved. A graph will be quite sufficient to convince most people that

TABLE 14.5

LETHAL DOSE (CODED BY SUBTRACTING 50 UNITS) OF UNITED STATES STANDARD OUABAIN, SLOW INTRAVENOUS INJECTION IN CAT TO ENDPOINT OF CARDIAC STOPPAGE

Rate of injection, $X \frac{\text{mg./kg./min.}}{1000}$					Total
1.04575	2.0915	4.183	8.366		
5	3	34	51		
9	6	34	56		
11	22	38	62		
13	27	40	63		
14	27	46	70		
16	28	58	73		
17	28	60	76		
20	37	60	89		
22	40	65	92		
28	42				
31	50				
31					
<i>SY</i>	217	310	435	632	1,594
Number cats	12	11	9	9	41
$\bar{y}$	18.1	28.2	48.3	70.2	38.9
<i>SY</i> ²	4,727	10,788	22,261	45,940	83,716

there is no evidence of curvilinearity. However, we shall make the test of significance as an illustration of the method.

First, the total sum of squares of the lethal doses is analyzed in the familiar manner of table 14.6. It is clear that the chief source of variation is that induced by the rates of injection—the group means differ significantly. We are interested in learning whether the trend of the means departs significantly from linearity.

For regression, Sx^2 and Sxy are required, corresponding to $Sy^2 = 21,744$

TABLE 14.6

ANALYSIS OF VARIANCE OF LETHAL DOSES OF OUABAIN AT FOUR RATES OF INJECTION

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Within rate groups (error)	37	5,651	152.7
Rates of injection	3	16,093	5,364**
Total	40	21,744	

in the table. The calculations differ somewhat from usual because of the occurrence of Y in groups. Thus,

$$SX = 12(1.04575) + 11(2.0915) + 9(4.183) + 9(8.366) = 148.496,$$

a *weighted* sum. Similarly,

$$SX^2 = 12(1.04575)^2 + \dots + 9(8.366)^2 = 848.628$$

From these is computed,

$$Sx^2 = SX^2 - (SX)^2/n = 848.628 - 537.831 = 310.797$$

For the sum of products of the observations,

$$SXY = (1.04575)(217) + (2.0915)(310) + (4.183)(435) + (8.366)(632) \\ = 7,982.21$$

Deducting the correction term, $(148.496)(1,594)/41 = 5,773.23$, we have,

$$S_{xy} = 7,982.21 - 5,773.23 = 2,208.98$$

Finally, the sum of squares for group means is copied from table 14.6 into table 14.7, then divided into the two parts, one attributed to linear regression and the other to deviations therefrom. The former is, as usual, $(S_{xy})^2/(Sx^2) = (2,208.98)^2/310.797 = 15,700$, having one degree of freedom. The remainder, 393, with two degrees of freedom furnishes the mean square corresponding to deviations from linear regression. This mean square is little greater than that for error. No more than random sampling departure from linear regression is indicated.

TABLE 14.7
ANALYSIS OF SUM OF SQUARES FOR GROUP MEANS

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Rates of injection (table 14.6)	3	16,093	
Linear regression, $(S_{xy})^2/(Sx^2)$	1	15,700	
Deviations from linear regression	2	393	196
Error (table 14.6)	37		153

$$F = 196/153 = 1.3$$

14.5—Test of departure from linear regression in covariance analysis. As in any other correlation and regression work, it is necessary in covariance to be assured that regression is linear, else modifications must be made in the methods. In some cases, randomized blocks for example, it is only the error regression that matters. If the error regression is rectified, the adjusted means are appropriate estimates and the test of significance is valid.

EXAMPLE 14.11—There is available some additional information about the chinch bugs in example 14.9. Each of the eight mortality percentages was an average of five subsamples of 100 bugs. The mortality percentages in the subsample are as follows:

Sub-sample of 100	Days Exposure							
	1	3	3.5	8	13.5	16	21	28
1	0	3	4	9	21	46	68	93
2	0	4	5	21	23	39	70	100
3	1	2	6	8	27	43	66	92
4	1	5	7	10	25	47	61	86
5	2	4	7	10	18	48	73	89
Total	4	18	29	58	114	223	338	460

Analyze the variance:

Within dates	421.2,	32 <i>df.</i> ,	mean square	13.2
Between dates	39,934.4,	7 <i>df.</i> ,	mean square	5,704.9

Although some heterogeneity is indicated, the mean square, 13.2, may be taken as the estimate of error against which to test any of the mean squares in example 14.9. For example, $F = 32.2/13.2 = 2.44$, $df. = 5$ and 32, shows that the departure of these data from second degree regression is not significant.

EXAMPLE 14.12—Swanson and Smith (15) determined the total nitrogen content (Y grams per 100 cc. of plasma) of rat blood plasma at nine ages (X days).

Number of Rats	Age								
	25	37	50	60	80	100	130	180	360
1	0.83	0.98	1.07	1.09	0.97	1.14	1.22	1.20	1.16
2	0.77	0.84	1.01	1.03	1.08	1.04	1.07	1.19	1.29
3	0.88	0.99	1.06	1.06	1.16	1.00	1.09	1.33	1.25
4	0.94	0.87	0.96	1.08	1.11	1.08	1.15	1.21	1.43
5	0.89	0.90	0.88	0.94	1.03	0.89	1.14	1.20	1.20
6	0.83	0.82	1.01	1.01	1.17	1.03	1.19	1.07	1.06
7	0.84	0.95	0.93		0.98	1.08	1.19	1.13	1.29
8	0.75	0.92	1.07		0.99	0.98	1.14	1.12	1.25
9	0.67	0.87	1.03			0.98	1.13		1.23
10	0.70		1.13			1.10	1.06		1.22
11	0.77		1.05				1.11		1.17
12	0.76		0.96				1.04		
13	0.75		1.01				1.29		
14	0.78		1.01				1.09		
15	0.76		0.94				1.14		
16	0.86		1.03				1.12		
17	0.78		1.06				1.29		
18	0.84		1.08				1.28		
19	0.85		1.09				1.10		
20	0.83		0.93				1.07		
21	0.85		0.95				1.10		
22	0.85								
23	0.83								
24	0.83								
25	0.70								

Verify the following:

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Total	118	2.8686	
Ages	8	2.3076	
Individuals in age groups (error)	110	0.5610	0.0051
Ages	8	2.3076	
Linear regression	1	1.4851	
Deviations	7	0.8225	0.1175**

For illustration we shall return to the sugar beet data of table 12.13. An agronomist, studying the findings of the covariance analysis, expressed surprise that the differences among the treatment mean weights per root were not significant. One phase of the question was developed in example 12.10. Another leads to a test of the linearity of the regression. The method is the same as that of section 14.3. The squares of the numbers of beets in the plots, X^2 , considered as a second independent variate, are used with the first powers, X , and the dependent yield, Y , in the multiple covariance analysis of table 14.8. The squares are expressed in units of 1,000, $(183)^2 = 33,489$ being entered in the data sheet as $X^2 = 33$. Corrections for origin are made in the usual fashion (section 13.10). In section 12.7 some of the preliminary calculations are described in detail.

TABLE 14.8
ANALYSIS OF VARIANCE AND COVARIANCE OF NUMBERS OF BEETS (X), THE SQUARE (X^2) AND YIELD (Y)

Source of Variation	Variable	X	X^2	Y
Total	X	152,158.00	88,165.00	4,163.69
Blocks		7,472.57	4,542.00	— 116.56
Treatments		116,020.33	66,953.67	3,598.05
Error		28,665.10	16,669.33	682.20
Treatment + error		144,685.43	83,623.00	4,280.25
Total	X^2		51,704.48	2,383.94
Blocks			2,782.19	— 71.88
Treatments			38,779.48	2,065.57
Error			10,142.81	390.25
Treatment + error			48,922.29	2,455.82
Total	Y			142.4022
Blocks				6.3134
Treatments				112.8562
Error				23.2326
Treatment + error				136.0888

The sums of squares and products for error lead to the multiple correlation, $R^2 = 0.7029$. The product of $Sy^2 = 23.2326$ (in the line for error) by $(1 - R^2) = 0.2971$ is the sum of squares of errors of estimate from quadratic regression, entered in table 14.9 along with a number

TABLE 14.9
TEST OF SIGNIFICANCE OF DEVIATIONS FROM *Error* LINEAR REGRESSION
Sugar Beets

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Deviations from linear regression, $Sy^2(1 - r^2)$	29	6.9969	0.2465
Deviations from quadratic regression, $Sy^2(1 - R^2)$	28	6.9024	
Reduction due to quadratic regression	1	0.0945	

from table 12.14 giving similar information about deviations from linear regression. The sum of squares ($df. = 1$) credited to curvilinearity of regression, 0.0945, is less than the error mean square. Certainly, departure from linearity of regression is negligible.

In some other example similar to this the regression might turn out to be significantly nonlinear. You would then wish to test the significance of the differences among means adjusted by the curvilinear regression. Turning back to table 14.8, you would make a second multiple regression of the *treatment + error* sums of squares and products, resulting in $R^2 = 0.9343$. The test of significance is then carried through (table 14.10) in the familiar way. In the present example, since the degree of freedom for second degree regression does not carry an average amount of sum of squares, the F -value is smaller than that of table 12.14. One would use this test only if there were a significant reduction in sum of squares attributable to curvilinearity of regression.

TABLE 14.10
TEST OF SIGNIFICANCE OF DIFFERENCES AMONG MEAN YIELDS ADJUSTED
FOR CURVILINEAR REGRESSION
Sugar Beets

Source of Variation	Sum of Squares Sy^2	$1 - R^2$	Errors of Estimate		
			Sum of Squares	Degrees of Freedom	Mean Square
Treatment + error	136.0888	0.0657	8.9410	34	0.2465
Error	23.2326	0.2971	6.9024	28	
Treatment			2.0386	6	0.3398

$$F = 0.3398/0.2465 = 1.38$$

EXAMPLE 14.13—The test for curvilinearity in multiple regression is similar to those in the last two sections. Emmert (4) measured the nitrogen, N , and phosphorous, P , (parts per million) in the stems of potato plants on June 15, 1935, then recorded the associated yield in pounds, Y , from each of 37 plots. The multiple regression of Y on N and P had $R = 0.6186$. Since there was some evidence of curvilinearity, the variables N^2 and P^2 were added. The multiple regression of Y on N , P , N^2 and P^2 yielded $R = 0.6522$. The increase in R indicates some departure from linearity. The b -primes showed that this was due to the term in P^2 only, the N^2 term yielding no additional information. The latter term being dropped, therefore, the regression of Y on N , P and P^2 gave $R = 0.6522$ as before. Since only a test of significance is desired, it isn't necessary to multiply each $(1 - R^2)$ by Sy^2 . The test is as follows:

Source of Variation	Degrees of Freedom	$1 - R^2$	Factor for Test of Significance
Regression of Y on N and P	34	0.6174	
Regression of Y on N , P and P^2	33	0.5747	0.0174
Curvilinearity	1	0.0427	0.0427

$$F = 0.0427/0.0174 = 2.45, \quad df. = 1 \text{ and } 33$$

F being far short of its 5% value, the departure from linearity is attributable to accidents of sampling.

The data are below.

N	P	P^2	Y	N	P	P^2	Y
139	195	38	33	330	260	68	46
143	146	21	43	380	134	18	60
147	120	14	40	436	130	17	57
154	178	32	49	436	200	40	64
162	266	71	43	470	134	18	55
182	200	40	27	476	160	26	54
186	260	68	55	555	155	24	37
200	240	58	51	556	188	35	42
208	170	29	45	626	125	16	45
213	195	38	41	700	146	21	63
233	146	21	57	750	178	32	52
257	160	26	38	834	188	35	46
264	146	21	48	834	178	32	56
264	146	21	39	908	105	11	65
264	160	26	40	1125	160	26	46
264	250	62	38	1200	122	15	58
265	175	31	52	1250	140	20	80
286	114	13	49	1625	150	22	70
312	155	24	51				

$$Y = 0.0170N - 0.440P + 1.10P^2 + 82.5$$

14.6—The fitting of orthogonal polynomials. A special method has been perfected by Fisher (5) for fitting the polynomial,

$$Y = a + bX + cX^2 + dX^3 + \dots$$

This is a flexible curve, adaptable to fitting a wide assortment of data. It includes the straight line and the parabola as special cases. Fisher's

method is presented both for its ready utility and for its theoretical value. We shall illustrate only the case in which the values of X are spaced at unit intervals, each accompanied by a single value of Y . If X occurs in equally-spaced intervals different from unity, it may be coded to unit intervals by dividing each value by the common interval. In data having X unequally spaced or having more than one Y for some or all of the X 's, use the method of section 14.4.

One advantage of the present method is that the fitting can be carried through conveniently in successive stages, the success of fitting terms of higher and higher degree being observed and tested for significance at each stage. The fitting will be terminated with the fourth degree term in the example below. Those who wish formulas for higher degree terms may infer them readily with the assistance of the footnotes. Professor Fisher gives general formulas in his book.

For illustration, the chick embryo data of table 14.1 will be used. The dry weights are copied in the first column of table 14.11. The sum,

TABLE 14.11
SUMS AND DERIVED QUANTITIES FOR FITTING A POLYNOMIAL TO CHICK EMBRYO DATA

Y	2	3	4	5
0.029	0.029	0.029	0.029	0.029
0.052	0.081	0.110	0.139	0.168
0.079	0.160	0.270	0.409	0.577
0.125	0.285	0.555	0.964	1.541
0.181	0.466	1.021	1.985	3.526
0.261	0.727	1.748	3.733	7.259
0.425	1.152	2.900	6.633	13.892
0.738	1.890	4.790	11.423	25.315
1.130	3.020	7.810	19.233	44.548
1.882	4.902	12.712	31.945	76.493
2.812	7.714	20.426	52.371	128.864
$S_1 = 7.714 \quad S_2 = 20.426 \quad S_3 = 52.371 \quad S_4 = 128.864 \quad S_5 = 302.212$				
$a = 0.701273 \quad b = 0.309485 \quad c = 0.183115 \quad d = 0.128735 \quad e = 0.100637$				
$a' = 0.701273 \quad b' = 0.391788 \quad c' = 0.139048 \quad d' = 0.031838 \quad e' = 0.003066$				
$A = 0.701273 \quad B = 0.235073 \quad C = 0.0463493 \quad D = 0.00619072 \quad E = 0.00038325$				

$S_1 = 7.714$, is divided by $n = 11$, the number of items in the set, to get the mean, $a = 0.701273$, as specified in the first of the following formulas:

$$a = \frac{S_1}{n} = \bar{y}$$

$$b = \frac{(1)(2)}{n(n+1)} (S_2)$$

$$c = \frac{(1)(2)(3)}{n(n+1)(n+2)} (S_3) \quad .$$

$$d = \frac{(1)(2)(3)(4)}{n(n+1)(n+2)(n+3)} (S_4)$$

$$e = \frac{(1)(2)(3)(4)(5)}{n(n+1)(n+2)(n+3)(n+4)} (S_5)$$

The remaining formulas will be used later.

The mean, a , is copied in the next line as a' , the two being equal according to the first of the next set of formulas:

$$a' = a$$

$$b' = a - b$$

$$c' = a - 3b + 2c$$

$$d' = a - 6b + 10c - 5d$$

$$e' = a - 10b + 30c - 35d + 14e^*$$

Again, $A = a' = a$ in the first of the third set:

$$A = a'$$

$$B = \frac{6}{n-1} (b')$$

$$C = \frac{30}{(n-1)(n-2)} (c')$$

$$D = \frac{140}{(n-1)(n-2)(n-3)} (d')$$

$$E = \frac{630^\dagger}{(n-1)(n-2)(n-3)(n-4)} (e')$$

Finally, the sum of squares of deviations from mean is got by subtracting from $SY^2 = 13.577730$ the correction term $nA^2 = (11)(0.701273)^2 = 5.409622$, set down as the first of these quantities:

$$nA^2 = \frac{(SY)^2}{n}$$

$$\frac{n(n^2-1)}{12} (B^2)$$

* The next three of these equations are

$$f' = a - 15b + 70c - 140d + 126e - 42f$$

$$g' = a - 21b + 140c - 420d + 630e - 462f + 132g$$

$$h' = a - 28b + 252c - 1050d + 2310e - 2772f + 1716g - 429h$$

† The numerators of the next three terms are 2,772, 12,012, 51,480.

$$\frac{n(n^2 - 1)(n^2 - 4)}{180} (C^2)$$

$$\frac{n(n^2 - 1)(n^2 - 4)(n^2 - 9)}{2,800} (D^2)$$

$$\frac{n(n^2 - 1)(n^2 - 4)(n^2 - 9)(n^2 - 16)}{44,100^*} (E^2)$$

The remainder, $S_y^2 = SY^2 - nA^2 = 8.168108$, is the end of the first stage in the fitting. A horizontal straight line, $Y = 0.701273$, has been fitted to the data. The sum of squares of deviations from this line is S_y^2 .

The second stage starts with the sums in column 2 of the table. After the first item is copied at the top of the column, the successive sums are:

$$0.029 + 0.052 = 0.081$$

$$0.081 + 0.079 = 0.160$$

$$0.160 + 0.125 = 0.285, \text{ etc.}$$

These are readily run up on an adding machine. The last sum, 7.714, is the same as S_1 , a convenient check. The sum of column 2 is $S_2 = 20.426$.

Going through the sets of formulas again, there result:

$$b = \frac{(1)(2)}{(11)(12)} (20.426) = 0.309485$$

$$b' = 0.701273 - 0.309485 = 0.391788$$

$$B = \frac{6}{10} (0.391788) = 0.235073$$

$$\begin{aligned} S(Y - \hat{Y})^2 &= 8.168108 - \frac{11(11^2 - 1)}{12} (0.235073)^2 \\ &= 8.168108 - 6.078525 = 2.089583 \end{aligned}$$

The reduction in sum of squares due to fitting the linear regression is 6.078525 with 1 degree of freedom. This may be tested for significance by comparing it with the remainder, 2.089583, having 9 *d.f.* The mean square for error is, therefore, $2.089583/9 = 0.232$, yielding $F = 6.079/0.232 = 26.2$, a highly significant value for *d.f.* = 1 and 9.

* The denominators of the next three terms are 698,544, 11,099,088, 176,679,360.

The third stage involves fitting the second degree term. Starting with $S_2 = 52.371$,

$$c = \frac{(1)(2)(3)}{(11)(12)(13)} (52.371) = 0.183115$$

$$c' = (0.701273) - 3(0.309485) + 2(0.183115) = 0.139048$$

$$C = \frac{30}{(10)(9)} (0.139048) = 0.0463493$$

$$\begin{aligned} S(Y - \hat{Y})^2 &= 2.089583 - \frac{(11)(11^2 - 1)(11^2 - 4)}{180} (0.0463493)^2 \\ &= 2.089583 - 1.843205 = 0.246378 \end{aligned}$$

The mean square for testing is now $0.246378/8 = 0.0308$, so that $F = 1.843/0.0309 = 59.8$, which with $d.f. = 1$ and 8 is highly significant.

In the next stage, the third degree term is fitted with this result:

$$S(Y - \hat{Y})^2 = 0.246378 - 0.236757 = 0.009621$$

This yields $F = 0.237/0.00137$, $d.f. = 1$ and 7 , another highly significant reduction.

One would be justified in stopping at this point. Since $1 - R^2$ is the ratio of the remaining sum of squares to Sy^2 , we have

$$\begin{aligned} 1 - R^2 &= 0.009621/8.1681 = 0.001178 \\ R &= \sqrt{0.998822} = 0.9994, \end{aligned}$$

about as high a correlation between estimated and actual values as one could wish. However, for two reasons the fitting will be carried one stage further: to illustrate the method, and because the third degree polynomial exhibits certain snakelike curves that make it unsuited to the chick embryo data.

Taking the last of the successive sets of formulas, with $S_3 = 302.212$, one arrives at

$$S(Y - \hat{Y})^2 = 0.009621 - 0.006049 = 0.003572,$$

so that $F = 0.006049/0.000595 = 10.2$, a significant, but not highly significant, reduction.

For many purposes, the job is complete. Some, though, may wish to plot the curves, and some may require the equations. Those who need the graphs but not the equations should follow the procedure of the next few paragraphs. If equations are necessary, skip to the sixth paragraph below.

In table 14.12 are entered some polynomials with corresponding coefficients which enable one to build up the desired values of $\hat{Y}$ by adding sets of differences to a *terminal value*. The coefficients depend on the size of

TABLE 14.12
COEFFICIENTS AND POLYNOMIALS FOR TERMINAL VALUES AND DIFFERENCES FOR FITTING
TERMS UP TO THE SEVENTH DEGREE

Degree of Polynomial	Coefficient Depending on Sample Size	Polynomial Ending With Term of Degree							
		0	1	2	3	4	5	6	7
0	(Terminal Value)	$a' + 3b' + 5c' + 7d' + 9e' + 11f' + 13g' + 15h'$							
1	$-\frac{6}{n-1}$	$b' + 5c' + 14d' + 30e' + 55f' + 91g' + 140h'$							
2	$\frac{60}{(n-1)(n-2)}$	$c' + 7d' + 27e' + 77f' + 182g' + 378h'$							
3	$\frac{-840}{(n-1)(n-2)(n-3)}$	$d' + 9e' + 44f' + 156g' + 450h'$							
4	$\frac{15,120}{(n-1) \dots (n-4)}$	$e' + 11f' + 65g' + 275h'$							
5	$\frac{-332,640}{(n-1) \dots (n-5)}$	$f' + 13g' + 90h'$							
6	$\frac{8,648,640}{(n-1) \dots (n-6)}$	$g' + 15h'$							
7	$\frac{-259,459,200}{(n-1) \dots (n-7)}$	h'							

the sample, while the polynomials are determined by the degree of the last term fitted. As a first illustration of the method we shall calculate points on the best fitting straight line.

Since the last term now to be considered is of degree one, the terminal value is $(a' + 3b') = 0.701273 + 3(0.391788) = 1.8766$. The *first difference* is

$$-\frac{6}{(n-1)}(b') = -\frac{6}{(11-1)}(0.391788) = -0.2351$$

By successive additions of this first difference to the terminal value, 1.8766, one gets the successive values of Y , beginning with the largest corresponding to $X = 16$:

$$\begin{aligned} 1.8766 + (-0.2351) &= 1.6415 \\ 1.6415 + (-0.2351) &= 1.4064 \end{aligned}$$

X	Polynomial Value, Y	First Difference
6	-0.474	
7	-0.239	
8	-0.004	
9	0.231	
10	0.466	
11	0.701	
12	0.936	
13	1.171	
14	1.406	
15	1.642	
16	Terminal value = 1.8766	-0.2351

In the calculating machine all the four decimals are carried, but the last one is deleted in the column of polynomial values. These values are plotted in figure 14.3. The straight line is a poor fit, of course, but it does pass through the point $(\bar{x}, \bar{y})$; that is, (11, 0.701).

Next, let us calculate the ordinates for the best fitting parabola. From the table, ending with degree 2:

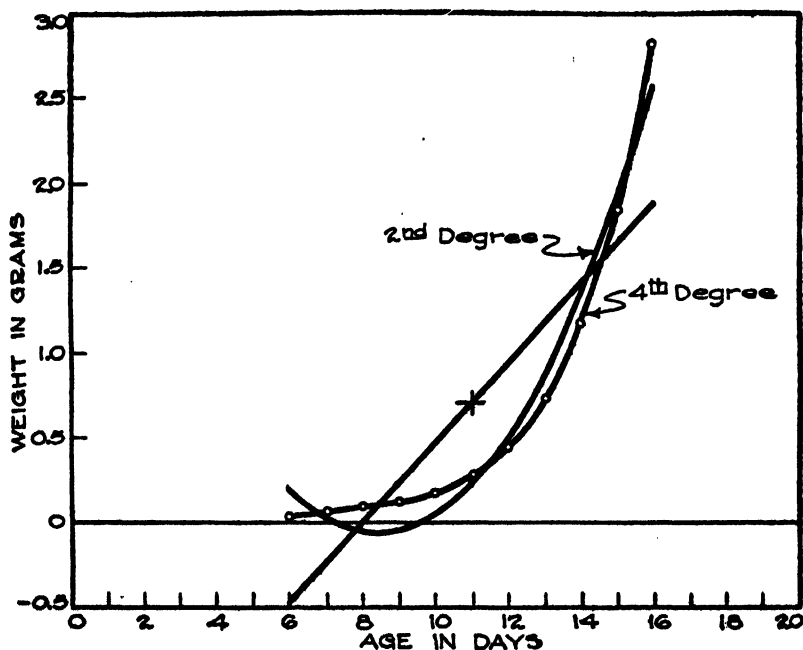


FIG. 14.3—Graphs of polynomials of first, second, and fourth degree fitted to chick embryo data of table 14.1.

$$\text{Terminal value} = a' + 3b' + 5c'$$

$$= 0.701273 + 3(0.391788) + 5(0.139048) = 2.57188$$

$$\text{First difference} = -\frac{6}{n-1} (b' + 5c')$$

$$= -\frac{6}{10} [0.391788 + 5(0.139048)] = -0.65222$$

$$\text{Second difference} = \frac{60}{(n-1)(n-2)} (c')$$

$$= \frac{60}{(10)(9)} (0.139048) = 0.092699$$

The calculations are carried through like this:

X	Y	First Difference	Second Difference
6	0.222	0.2748	
7	0.039	0.1821	
8	-0.050	0.0894	
9	-0.047	-0.0033	
10	0.049	-0.0960	
11	0.238	-0.1887	
12	0.519	-0.2814	
13	0.893	-0.3741	
14	1.360	-0.4668	
15	1.920	-0.5595	
16	Terminal value = 2.57188	-0.65222	0.092699

The second difference is added successively to the bottom first difference,

$$0.092699 + (-0.65222) = -0.5595,$$

and to each of the sums above it. Each first difference, starting with the last, is added to a value of Y to get the one next above. Thus,

$$-0.65222 + 2.57188 = 1.920$$

$$-0.5595 + 1.920 = 1.360, \text{ etc.}$$

Y is plotted against X to make the graph of the parabola.

Passing to the fourth degree polynomial, the terminal value and differences are:

$$\text{Terminal value} = a' + 3b' + 5c' + 7d' + 9e' = 2.822337$$

$$\text{First difference} = -\frac{6}{n-1} (b' + 5c' + 14d' + 30e') = -0.974844$$

$$\text{Second difference} = \frac{60}{(n-1)(n-2)} (c' + 7d' + 27e') = 0.296464$$

$$\text{Third difference} = \frac{-840}{(n-1)(n-2)(n-3)} (d' + 9e') = -0.069337$$

$$\text{Fourth difference} = \frac{15,120}{(n-1)(n-2)(n-3)(n-4)} (e') = 0.009198$$

These are entered in the last line of table 14.13. The third differences

TABLE 14.13
CALCULATION OF THE FOURTH DEGREE POLYNOMIAL VALUES FOR CHICK EMBRYO DATA

<i>X</i>	<i>Y</i>	First Difference	Second Difference	Third Difference	Fourth Difference
6	0.026				
7	0.056	-0.0302			
8	0.086	-0.0295	-0.00069		
9	0.119	-0.0337	0.00426	-0.004951	
10	0.171	-0.0521	0.01841	-0.014149	
11	0.265	-0.0939	0.04176	-0.023347	
12	0.434	-0.1682	0.07430	-0.032545	
13	0.718	-0.2843	0.11605	-0.041743	
14	1.169	-0.4512	0.16699	-0.050941	
15	1.847	-0.6784	0.22713	-0.060139	
16	2.8223	-0.97484	0.296464	-0.069337	0.009198

result from successive additions of 0.009198. The second differences are calculated thus:

$$-0.069337 + 0.296464 = 0.22713$$

$$-0.060139 + 0.22713 = 0.16699, \text{ etc.}$$

All the significant figures are carried in the machine, as before, but one is dropped in copying. The first differences are got in a similar manner, then the polynomial values. These last, plotted on the graph, specify the fourth degree curve. It passes so nearly through the data points that the discrepancies cannot be depicted.

The information now at hand is all that is needed in most cases. The goodness of fit is determined both by *R* and tests of significance of the reductions in sum of squares. The graph not only gives visual representation but serves for interpolation accurately enough for ordinary purposes. There would seem to be little incentive to derive the equation of the curve.

If the equation is desired, omit the computation of the polynomial values in the preceding few paragraphs. After the equation is set up as

described below, points on the graph can be computed at will. The equation is of the form,

$$Y = A + BX_1 + CX_2 + \dots$$

where $A, B, C, \dots$ are the quantities last calculated in table 14.11, and X_1, X_2 , etc., are themselves polynomials in X . It is in this part of the computation that X is assumed to occur at unit intervals. If the actual intervals are equal but not unity, as for example 30 days or 5 years, they may be coded easily by division. Practically, they are simply rewritten as 1, 2, 3, . . . , the new unit being 30 days or 5 years.

One last set of formulas is needed:

$$X_1 = X - \bar{x}$$

$$X_2 = X_1^2 - \frac{n^2 - 1}{12}$$

$$X_3 = X_1^3 - \frac{3n^2 - 7}{20} (X_1)$$

$$X_4 = X_1^4 - \frac{3n^2 - 13}{14} (X_1^2) + \frac{3(n^2 - 1)(n^2 - 9)}{560}$$

$$X_5 = X_1^5 - \frac{5(n^2 - 7)}{18} (X_1^3) + \frac{15n^4 - 230n^2 + 407}{1,008} (X_1)$$

From table 14.1, $\bar{x} = 11$ days and $n = 11$. Hence,

$$X_1 = X - 11$$

$$X_2 = (X - 11)^2 - 10 = X^2 - 22X + 111$$

$$X_3 = (X - 11)^3 - 17.8(X - 11) = X^3 - 33X^2 + 345.2X - 1,135.2$$

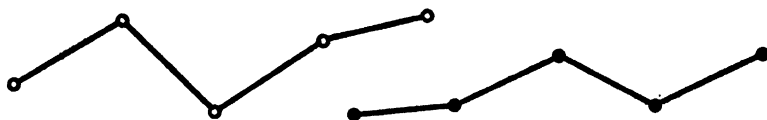
$$\begin{aligned} X_4 &= (X - 11)^4 - 25(X - 11)^2 + 72 \\ &= X^4 - 44X^3 + 701X^2 - 4,774X + 11,688 \end{aligned}$$

Any equation up to the fourth degree may be written from the data now before us. As examples, the equation of the best fitting straight line is, $Y = A + BX_1 = 0.701273 + 0.235073(X - 11) = 0.2351X - 1.885$, while that of the fourth degree polynomial is,

$$\begin{aligned} Y &= A + BX_1 + CX_2 + DX_3 + EX_4 \\ &= 0.701273 + 0.235073(X - 11) + 0.046349(X^2 - 22X \\ &\quad + 111) + 0.0061907(X^3 - 33X^2 + 345.2X - 1,135.2) \\ &\quad + 0.00038325(X^4 - 44X^3 + 701X^2 - 4,774X + 11,688) \\ &= 0.712 - 0.4772X + 0.11071X^2 - 0.010672X^3 + 0.0003832X^4 \end{aligned}$$

By substituting the several values of X days in this equation, the polynomial values may be calculated and plotted for the graph.

The student well may question the advisability of fitting curves. A stupendous amount of time has been wasted in ill-advised curve fitting. Only when the end in view is clear should the task be undertaken. Often a graph of the data points is sufficient. Represent them by small circles or heavy dots. If desired, they may be connected by light line segments.



Avoid drawing "eye-fitted" curves. They are highly subjective and are apt to be misleading to both the perpetrator and the victim. Interpolation with these links stands a better chance of being good than does estimating by means of even the most artistic curves. If you know enough about the law to say that there ought to be a curve, then you should be able to write an equation expressing the law and fit it to the data. If you wish to evaluate a correlation when regression is nonlinear, or if you wish to test significance, then suitable curved regressions must be fitted. Occasionally, fitted curves are required for interpolation. In many cases, graphical representation of the data is sufficient.

EXAMPLE 14.14—The following five points lie on the parabola, $Y = 9 + 2X - X^2$: (0, 9), (1, 10), (2, 9), (3, 6), and (4, 1). Using Fisher's method, show that the sum of squares remaining after fitting the second degree term is zero. Calculate the polynomial values. These will be identical with the pairs of values given above. When you plot them, the points will lie exactly on the parabola whose equation is given. Derive this equation by the use of the formulas up to the second degree.

EXAMPLE 14.15—Here are some points on the cubic, $Y = 9X - 6X^2 + X^3$: (0, 0), (1, 4), (2, 2), (3, 0), (4, 4), (5, 20). Carry through all the computations by Fisher's method, including the best fitting straight line and parabola. There will be no remaining sum of squares after fitting the third degree term and the polynomial values at that stage of the fitting will be exactly those you start with. Plot the best fitting line and the best fitting parabola and derive their equations.

EXAMPLE 14.16—Reed and Holland (13) reported the following average heights of sunflowers:

Week	1	2	3	4	5	6	7	8	9	10	11	12
Height	18	36	68	98	131	170	206	228	247	250	254	254
Polynomial values	15.0	38.9	68.0	100.6	134.6	168.0	198.9	225.5	245.6	257.5	259.1	248.4

The polynomial values are those for the cubic. Verify and plot. The remaining sum of squares after fitting the cubic is 211.03. Calculate the equation of the polynomial, then verify by duplicating the polynomial values given above.

EXAMPLE 14.17—The fit of the cubic to the sunflower heights is not entirely satisfactory, since it turns downward in the twelfth week. A great deal has been written

about a more suitable curve, the *logistic*, or the curve of autocatalytic growth. An excellent account, including bibliography, is given by Pearl (11).

EXAMPLE 14.18—Those who deal with dosage-mortality and time-mortality curves will find an excellent treatment in three articles by Bliss (1).

REFERENCES

1. C. I. BLISS. *Annals of Applied Biology*, 22:134 and 22:307 (1935); 24:815 (1937).
2. GEORGE C. DECKER and FLOYD ANDRE. *Iowa State College Journal of Science*, 10:403 (1936).
3. W. EDWARDS DEMING. *Statistical Adjustment of Data*. John Wiley and Sons, Inc., New York (1943).
4. E. M. EMMERT. *Journal of the American Society of Agronomy*, 29:213 (1937).
5. R. A. FISHER. *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh (1925), section 27.
6. W. F. GEDDES. *Canadian Journal of Research*, 1:528 (1929).
7. JOSEPH D. GRANT. *Finite Differences*. Edwards Bros., Ann Arbor (1926).
8. A. C. HASKELL. *How to make and use Graphical Charts*. Codex Book Co., Inc., New York (1919).
9. JOSEPH LIPKA. *Graphical and Mechanical Computation*. John Wiley and Sons, Inc., New York (1918).
10. W. H. METZGER. *Journal of the American Society of Agronomy*, 27:653 (1935).
11. RAYMOND PEARL. *Medical Biometry and Statistics*. W. B. Saunders Co., Philadelphia, 2nd Edition (1930).
12. ROBERT PENQUITE. Thesis submitted for the degree Doctor of Philosophy, Iowa State College (1936).
13. H. S. REED and R. H. HOLLAND. *Proceedings of the National Academy of Science*, 5:140 (1919).
14. T. R. RUNNING. *Empirical Formulas*. John Wiley and Sons, Inc., New York (1917).
15. PEARL P. SWANSON and A. H. SMITH. *The Journal of Biological Chemistry* 97:745 (1932).
16. O. W. WILLCOX. *Principles of Agrobiolgy*. Palmer Publishing Corporation, New York (1930).

Chapter 15

INDIVIDUAL DEGREES OF FREEDOM

15.1—Introduction. Mean squares for single degrees of freedom were implied in chapters 2 and 4, where two means were compared; and in chapter 6, where the regression coefficient was studied. The F -test was applied to corresponding null hypotheses in chapters 10 and 11. Furthermore, in the latter two chapters the subdivision of degrees of freedom for treatments led to the evaluation of one-degree comparisons (section 11.6, for example). Some extension and simplification of these earlier methods will be made in the present chapter.

Experiments may be designed, often rather easily, for individual comparisons among the treatments. If so, the sum of squares for n treatments may be subdivided into $n-1$ parts, each corresponding to a single degree of freedom; and these parts may be tested individually against experimental error. The result may be a notable increase in the information furnished by the experiment. Several such experiments will be presented.

15.2—A short method of calculating mean squares. Several times already we have had occasion to apply a special case of this short method. An instance occurred in section 10.5 where the mean square for the difference between the means of two lots of hormone-injected chicks was got from the sums in this way:

$$\frac{(1,067 - 616)^2}{22} = 9,245.5$$

It should be fixed in mind that this method applies only if the two sums are made up of the same numbers of individuals; that is, $k_1 = k_2 = k$. It is informative, also, to know that the divisor has two factors, the formula being

$$\frac{(S_1 - S_2)^2}{2k}$$

The common size of the two groups is k , whereas the factor, 2, is a special case of a theorem to be used below: 2 is a consequence of the fact that the variance of differences is twice that of the corresponding individuals (section 11.5 contains an application).

Even if there are more than two treatments, experiments are often designed for individual comparisons. An example is found in section 10.12 where the concentration of carbon dioxide in soil air was studied. It is

clearly appropriate to compare the mean concentration in the treated plots,

$$(86 + 78)/2 = 82,$$

with that in the check plot, 99. The comparison is most readily made by use of the sum, $86 + 78 = 164$, instead of the mean. First, let us employ the familiar methods for calculating a sum of squares (table 10.14, fourth step). In each sum, k is the number of items involved. The computation set out in the table herewith leads to the sum of squares of deviations from

Number of Plots	Number of Observations	Sum of Observations	(Sum) ² / k
1	9	99	1,089.00
2	18	164	1,494.22
3	27	263	2,583.22
Correction = $(263)^2/27$			2,561.81
Sum of squares			21.41

mean, 21.41. For the test of significance, this mean square is compared with the error mean square, 48.82, in table 10.20. Since the figure for the present individual comparison is less than error, nonsignificance is obvious. Nevertheless, you will be struck by the large size of the mean square in this comparison as contrasted with the mean square for 2 *df.* in the table 10.20. Most of the variation lies between the treated plots and check. This feature will be examined further in the next section.

Now for the new method of computation. If you carry through the processes of the foregoing paragraph with literal symbols instead of the Arabic numerals, you will reach this formula for the sum of squares:

$$\frac{(2S_1 - S_2 - S_3)^2}{6k},$$

where $S_1 = 99$, $S_2 = 86$, $S_3 = 78$, and $k = 9$. By use of this formula the calculation is very easy:

$$\frac{[2(99) - 86 - 78]^2}{6(9)} = 21.41$$

Formulas for individual comparisons, such as this one, have led to a new manner of thinking about them. One soon learns to contrast the sum, $S_2 + S_3$, with twice the compared value, S_1 . Essentially this is the same as subtracting the mean, $(S_2 + S_3)/2$, from S_1 . What about the figure 6? You recall a similar figure 2, in the formula for two groups of

equal size. There is an easy rule for arriving at this factor in the denominator. First, write down the comparison you intend to make:

$$S_1 - S_2 \text{ or } 2S_1 - S_2 - S_3$$

Then add the squares of the coefficients of the letters:

$$1^2 + (-1)^2 = 2, \text{ or } (2)^2 + (-1)^2 + (-1)^2 = 6$$

You will find that this rule always leads to the correct value of the coefficient of k in the denominator.

A second example involving an experiment with more than two groups is found in example 11.17, where four seed treatments for soybeans were tried along with no treatment. The numbers of plants emerging in 5 blocks were,

Untreated		446
Arasan	469	
Spargon	459	
Semesan, Jr.	467	
Fermate	471	
Average of treated		466.5

Instead of comparing the untreated with the mean of the treated, the new scheme is to use a difference which is 4 times as great:

$$(469 + 459 + 467 + 471) - 4(446) = 1866 - 1784 = 82$$

Symbolically, this may be written,

$$S_2 + S_3 + S_4 + S_5 - 4S_1$$

the sum of the squares of the coefficients being

$$1^2 + 1^2 + 1^2 + 1^2 + 4^2 = 20$$

Since $k = 5$, the mean square is,

$$\frac{(82)^2}{(20)(5)} = 67.24$$

This result checks with that of the longer method.

Please notice that in every comparison the sum of the coefficients is zero. This reminds you of the theorem that the sum of the deviations from mean is zero. Also worthy of note is the fact that the quantity squared in the comparison numerator is always the difference between two sums.

EXAMPLE 15.1—For the chick-hormone data, the comb weights of the 2 lots (11 chicks each) averaged 97 and 56 mg. Calculate the mean square from the formula,

$$\frac{k(\bar{x}_1 - \bar{x}_2)^2}{2}$$

EXAMPLE 15.2—The mean numbers of plants emerging from the plots of soy

beans were $\bar{x}_1 = 446/5 = 89.2$, $\bar{x}_2 = 93.8$, $\bar{x}_3 = 91.8$, $\bar{x}_4 = 93.4$, and $\bar{x}_5 = 94.2$. Calculate the mean square from the formula,

$$\frac{k(\bar{x}_2 + \bar{x}_3 + \bar{x}_4 + \bar{x}_5 - 4\bar{x}_1)^2}{20}$$

Thus, if means instead of sums are used, k is a multiplier instead of a divisor. The sum of the squares of the coefficients remains in the denominator.

EXAMPLE 15.3—Each of the following sums is based on 5 items: $S_1 = 20$, $S_2 = 30$, $S_3 = 35$, and $S_4 = 40$. Calculate the mean squares for the comparisons:

1. $S_1 + S_2 - S_3 - S_4$. Answer 31.25
2. $3S_1 - S_2 - S_3 - S_4$. Answer 33.75

EXAMPLE 15.4—In the first comparison of the preceding example you will get the same answer if you consider $S_1 = 20 + 30 = 50$ and $S_2 = 35 + 40 = 75$ with $k = 10$ instead of 5, then compare S_1 with S_2 . Try it.

EXAMPLE 15.5—Each of the following is the sum of 4 items: $S_1 = 12$, $S_2 = 18$, $S_3 = 20$, $S_4 = 30$, and $S_5 = 32$. The mean square for the comparison,

$$2S_1 + 2S_2 + 2S_3 - 3S_4 - 3S_5,$$

will be found to be 61.63. Here you are comparing the mean of 3 groups with that of 2

EXAMPLE 15.6—Derive the two formulas for mean squares,

$$\frac{k(\bar{x}_1 - \bar{x}_2)^2}{2} \text{ and } \frac{(S_1 - S_2)^2}{2k}$$

where $\bar{x}_1 = S_1/k$ and $\bar{x}_2 = S_2/k$.

EXAMPLE 15.7—Derive the formula for the mean square,

$$(2S_1 - S_2 - S_3)^2/(6k)$$

15.3—Orthogonal sets of comparisons. Professor Fisher and his associates have successfully maintained the thesis, early announced, that for each degree of freedom an appropriate sum of squares can be segregated; also that the total sum of squares for any group of $n-1$ comparisons can be partitioned into $n-1$ parts, each fit for testing one of the set of independent individual comparisons.

For an illustration, let us return to the first part of the foregoing section. One of the two degrees of freedom for plots, table 10.20, has been furnished with a suitable sum of squares, 21.41. The other may be associated with the comparison of concentration of carbon dioxide in the air of the two treated plots. The two comparisons may be symbolized thus:

$$\begin{array}{ll} \text{between treated plots,} & S_2 - S_3 \\ \text{check vs. treated plots,} & 2S_1 - S_2 - S_3 \end{array}$$

In the more concise form, the comparisons are:

S_1	S_2	S_3
0	+1	-1
2	-1	-1

Notice two important facts. Not only is each sum of coefficients zero, analogous to the sum of deviations from mean; but the sum of the products of the corresponding coefficients is zero,

$$(0)(2) + (1)(-1) + (-1)(-1) = 0,$$

analogous to zero covariance. The absence of correlation between the comparisons is the reason they are designated *orthogonal*.

Completing the calculations, the comparison between the treated plots leads to the sum of squares,

$$\frac{(86 - 78)^2}{(1^2 + 1^2)(9)} = 3.56$$

Thus the difference between the means for the two treated plots is obviously not significant. Finally, the sums of squares for the two independent comparisons add to

$$3.56 + 21.41 = 24.97,$$

which is the sum of squares for the two degrees of freedom for plots in table 10.20. This addition theorem is the same as that characterizing sums of squares in analysis of variance. We are merely carrying to its conclusion the process begun in chapter 10, the matching of sums of squares with degrees of freedom.

You may ask the reason for choosing the particular set of individual comparisons presented above. Clearly other sets might be selected. The check plot might have been compared with one of the treated, and these in turn with the second treated. Theoretically, there is no limit to the number of sets of comparisons that may be devised. For practical purposes, however, the comparisons must be meaningful, ordinarily planned as part of the experiment. Extra amounts of information are available from experiments designed for the utilization of the individual comparisons. Let us examine one of them.

The cooperative experiment on the effect of fertilizer, example 11.29, is of the type known as *factorial*. This design will be discussed more at length in section 15.9. Here, we notice that the 3 degrees of freedom for treatment may be identified with the following comparisons, each sum being composed of 16 sampling units:

Treatment Sum (decagrams)	Check 291	<i>N</i> 313	<i>P</i> 312	<i>NP</i> 314	Differ- ence	Mean Square
<i>N</i> vs. no <i>N</i>	—	+	—	+	24	9.00
<i>P</i> vs. no <i>P</i>	—	—	+	+	22	7.56
Interaction	+	—	—	+	—20	6.25
Sum of squares						22.81

The comparisons are such that the sum of the coefficients in each is zero and that the sum of the products in each pair is zero. It is to be expected, then, that the sum of the three mean squares will be the same as the sum of squares, 22.81, in the original analysis of variance. The coefficients of interaction are each the product of those above in the two main effects. As may be seen from the signs, interaction compares the yields of plots having both fertilizers and neither with those having one kind only.

Each mean square may be tested against experimental error by use of F : since all are smaller than error, 31.78, no more than sampling variation is indicated.

As foreseen in section 10.4, you are now able to test as many separate null hypotheses as there are degrees of freedom for treatments. This discloses a fascinating prospect for the design and analysis of experiments. Only a glimpse can be given in this text, but some references will be found at the end of the chapter.

EXAMPLE 15.8—The rat experiment reported in example 11.32 allows further subdivisions of the degrees of freedom for treatment. A set of orthogonal comparisons follows:

Comparison	High Level			Low Level			Difference	Mean Square
	Beef 1000	Cereal 859	Pork 995	Beef 792	Cereal 839	Pork 787		
High vs. low	+	+	+	—	—	—	436	3,168.3
Beef vs. pork	+	0	—	+	0	—	10	2.5
Interaction with level	+	0	—	—	0	+	0	0.0
Animal vs. vegetable	+	—2	+	+	—2	+	178	264.0
Interaction with level	+	—2	+	—	+2	—	376	1,178.1
(Error mean square, 214.6, $d.f.$ = 54)								4,612.9

Here, the 2 $d.f.$ for source of food have been divided into 1 for "Beef vs. Pork," and 1 for "Animal vs. Vegetable." The sum of the mean squares, $264.0 + 2.5 = 266.5$, is the same as the sum of squares, "Source," in the former table—the mean square is $266.5/2 = 133.2$. A similar subdivision of the interaction has been made.

The significant interaction of "Animal vs. Vegetable" with level is worth looking at. Multiply the sums by the indicated coefficients and make this 2×2 table:

	Animal	Vegetable
High	1,995	1,718
Low	1,579	1,678

It is evident that the level of intake of the animal foods (beef and pork) affected the rat gains more than did that of the cereal.

EXAMPLE 15.9—Make a 2×2 table showing the interaction of level with source of food (animal or vegetable), but containing sums only, without multiplication by 2:

1,995	859
1,579	839

Using the methods of chapter 11, calculate the interaction from this table. You get the same result, of course, but with more labor.

15.4—Nonorthogonal sets of comparisons. Many experiments are planned for sets of comparisons which are not orthogonal. This must not be considered a fault if the experiment accomplishes its purpose. The sum of squares for the set of comparisons will not equal that between means of groups, and statements about probability may not be exact, but the comparisons are not lacking in interest. A few examples will suffice to illustrate the situation.

In the swine nutrition experiment of example 11.10, lot 1 was fed a standard corn belt ration. The plan was to compare the gains produced by this ration with those of each of the other lots. The three comparisons have the following sums of squares:

$$(14.64 - 11.95)^2/2(10) = 0.3618^{**}$$

$$(14.64 - 13.25)^2/2(10) = 0.0966$$

$$(14.64 - 16.62)^2/2(10) = 0.1960^{**}$$

The mean square for error with 27 *d.f.* is 0.0231. Only the second comparison is not significant, and its *F*-value is just short of the 5% point.

You will observe that the sum of the three mean squares, 0.6544, is considerably less than the sum of squares for rations, 1.1986. A discrepancy might have been anticipated by writing down symbolically the comparisons with the first lot:

1	2	3	4
+	—	0	0
+	0	—	0
+	0	0	—

Although the sum of the coefficients in any line is zero, the sum of the products in each pair of lines is different from zero. In the first pair, for example,

$$(1)(1) + (-1)(0) + (0)(-1) + (0)(0) = 1$$

The three comparisons are not independent, but they do furnish the information for which the experiment was planned.

Where there is such a standard ration, or a standard variety, as a basis for comparisons, it is usually easier to calculate the significant mean difference,

$$(t_{.05})(s_d) = 2.052\sqrt{2(0.0231)/10} = 0.140 \text{ pound per day}$$

You can see readily that if this difference between two means were divided by the standard error of the mean difference, the resulting *t* would be

exactly at the 5% level for 27 *df*. Hence, the intervals from $1.464 - 0.140 = 1.324$ down, and from $1.464 + 0.140 = 1.604$ up, include all means differing significantly from the standard. This method of using the *t*-test leads to the identical conclusions derived from the individual comparisons.

The experiment reported in table 11.13 was so designed as to compare each humus and fertilizer treatment with another similar in all respects save one. These contrasts with their mean squares are:

$$\text{Straw: } (230.2 - 156.7)^2 / (2)(12) = 225.09^{**}$$

$$\text{PO}_4: (160.2 - 156.7)^2 / (2)(12) = 0.51$$

$$\text{Lime: } (164.6 - 160.2)^2 / (2)(12) = 0.81$$

Notice once more the failure of the sum of the three nonorthogonal comparisons to be the same as the sum of squares for humus treatments in table 11.14. The three comparisons are,

Check	Straw	Straw + PO ₄	Straw + PO ₄ + lime
+	-	0	0
0	+	-	0
0	0	+	-

not fulfilling the second requirement for an orthogonal set.

One could easily make up an orthogonal set of comparisons contrasting each added treatment with the aggregate of those previously included:

Straw vs. check,	+	-	0	0
PO ₄ vs. first two,	+	+	-2	0
Lime vs. first three,	+	+	+	-3

It is questionable whether such comparisons, though satisfying the tests for orthogonality, have validity in this experiment. For instance, the third might result from the contrast of lime with PO₄ or with straw or with check or with any combination. One must be satisfied with the information furnished by those comparisons which are built into the structure of the experiment.

EXAMPLE 15.10—Calculate the mean squares for the set of three orthogonal comparisons just discussed. From table 11.14 the sum must be 306.24.

EXAMPLE 15.11—In example 11.10, suppose the ration fed to lot 2 were considered standard. Make the comparisons:

Comparison	Lot				Mean Square
	1	2	3	4	
1	-	+	0	0	0.3618 ^{**}
2	0	+	-	0	0.0845
3	0	+	0	-	1.0904 ^{**}
Sum					1.5367

Here, the sum is greater than that for rations in the original analysis.

EXAMPLE 15.12—There are some interesting comparisons in the sugar beet experiment of table 12.13. Six of the treatments fall into two groups differing only in respect of superphosphate. The treatment, $P + N$, cannot be used because there is no contrasting treatment with N alone. That is the defect mentioned in example 12.10. The other six treatments give comparison 1. The potash and nitrogen comparisons lose one plot each for similar reasons. The fourth comparison, the interaction of P with

Comparison	Treatment							Mean Square
	None	p	k	pk	np	nk	npk	
1. P	—	+	—	+	0	—	+	81.27**
2. K	—	—	+	+	—	0	+	1.80
3. N	0	—	—	—	+	+	+	5.25
4. PK	1	—	—	+	0	0	0	0.63
5. NP	0	0	+	—	0	—	+	1.19

K , must include only four treatments because there was none with N alone. In an interaction, the plots receiving none and both treatments are compared with those having either one or the other. The interaction of N with P involves four treatments all including K .

The comparisons are not orthogonal and the set is not complete. Perhaps you can find some others of interest. An experiment in which the three fertilizers are tried in all combinations will be outlined in section 15.9.

EXAMPLE 15.13—In the foregoing example, set up one of the interaction comparisons in a 2×2 table and evaluate it by the usual computations of analysis of variance.

15.5—Comparisons not planned before the experiment is performed. It is not uncommon to plan variety or feeding trials with no standard variety or ration. Among the resulting n means there are $n - 1$ independent comparisons that may be made, yet the means can be paired in $n(n - 1)/2$ different ways. Unless a set of independent comparisons was incorporated when the design was decided upon there is a tendency to select from the larger number certain comparisons which may reflect no more than peculiarities in the data themselves. Now, careful examination of results is the investigator's duty: he should put the means into arrays, or segregate those which are alike from those which plainly differ; then interpret what he finds in the light of known facts. But he should avoid applying tests of significance to his arrangements because they are not random.

It is necessary to distinguish sharply between experiments in which F is large and those whose means do not differ significantly. An instance of the former was discussed in section 10.4, but such pronounced groupings may be lacking. It is common to see an array of variety means having almost normal distribution. The test has indicated that they differ significantly, yet there is no way to learn which are from different populations and which are from a common one. It is safe to infer that the high means are not from the same population as the low, and usually it is satisfactory to select an upper fraction for continued trial. Some find satisfaction in calculating the significant difference and comparing it with

the array, but apparently no connection is established with the 5% probability—this probability applies only to *random pairs*, whereas the significant difference extends over an *interval* in an arrayed sample. For such a test the probability is not known. The available information in this type of experiment consists of the means, which are unbiased estimates; the F -test of the hypothesis that these means are drawn from a common population; and the fiducial interval. This knowledge is sufficient to lead to decisions and to point toward new trials that should be made.

If testing the null hypothesis has produced a nonsignificant F , it usually is not profitable to attempt interpretations of differences among the means. Certainly, any oddities should be reviewed and should perhaps be made the basis for future experimentation, but they cannot warrant exact statements about probability. Reexamination of your sampling in section 3.2 will likely reveal striking arrangements of means. I selected these 6 in successive samples:

33.4, 26.4, 29.3, 30.9, 22.5, 29.3

Analysis of variance of the 60 pig gains in these 6 lots resulted in the following mean squares:

Means	5	142.7	
Individuals	54	111.7	$F = 1.28$

If this were the result of an actual experiment, the investigator might be tempted to infer a population difference corresponding to the two means, 22.5 and 33.4 pounds. He might justify this by calculating,

$$\bar{d} = 10.9 \text{ lbs.}, s_{\bar{d}} = \sqrt{2(111.7)/10} = 4.73, t = 2.30,$$

with probability about 0.03. But he would be forgetting that he is testing, not a random difference but the range of 6 random means. The proper test is that of table 5.5:

$$\text{range}/s_{\bar{d}}^n = 2.30, P = 0.5,$$

(or exactly, $\text{range}/\sigma_{\bar{d}} = 10.9/4.47 = 2.44$). If the ratio, $\text{range}/s_{\bar{d}}$, were greater than the 5% point, 4.04, then in a real experiment there would be reason for careful survey of the circumstances, including this one: how did it come about that so pronounced a difference was not anticipated? Ordinarily, discovery of nonsignificance of F is the end-point of the statistical examination.

EXAMPLE 15.14—Examine the means in the fertilizer experiment of example 11.28. They fall into three groups: A , BCD , and EFG . Calculate the fiducial half-interval, 19.4, and show that the population mean B might lie in either of the more extreme groups. What recommendations would you make?

15.6—Individual comparisons in regression. The method of fitting orthogonal polynomials, described in section 14.6, led to individual comparisons associated with (i) linear regression, (ii) parabolic regression

fitted to deviations from linear, etc. Each mean square can be tested for significance independently.

If the values of X are equally spaced, the mean squares for the successive comparisons are easily calculated from tabulated sets of coefficients (8) (11). Those in table 15.1 are sufficient for our purposes. To

TABLE 15.1

COEFFICIENTS AND DIVISORS FOR SETS OF ORTHOGONAL COMPARISONS IN REGRESSION. X IS SPACED AT EQUAL INTERVALS. READ FROM TOP TO BOTTOM OF COLUMNS

Number of Group	Degree of Polynomial														
	1	1	2	1	2	3	1	2	3	4	1	2	3	4	5
1	-	-	+	-3	+	-	-2	+	-	+	-5	+	-5	+	-
2	+	0	-2	-	-	+3	-	-	+2	-4	-3	-	+7	-3	+5
3		+	+	+	-	-3	0	-2	0	+6	-	-4	+4	+2	-10
4				+3	+	+	+	-	-2	-4	+	-4	-4	+2	+10
5							+2	+2	+	+	+3	-	-7	-3	-5
6											+5	+5	+5	+	+
Divisor	2	2	6	20	4	20	10	14	10	70	70	84	180	28	252

illustrate their use, let us recalculate the mean square for linear regression of yield of millet on spacing, section 11.7. Since there are five spacings (groups), and since we wish only the linear or first degree regression, we copy the sums from table 11.11 and the coefficients from table 15.1:

$$-2(1,349) - 1(1,314) + 0(1,262) + 1(1,191) + 2(1,188)$$

That is, $-4,012 + 3,567 = -445$ is the difference for the numerator of the mean square. The sum of the squares of the coefficients is 10, and $k = 5$; therefore,

$$\frac{(-445)^2}{(10)(5)} = 3,960.5$$

is the mean square attributable to linear regression, the same result as the former except for rounding.

Linear regression may not account for the major portion of the sum of squares. An illustration is found in the seasonal yields of oranges at Riverside. Since the varieties of section 11.16 were only dummies, the data are combined into annual totals:

1924	1925	1926	1927
2,313	2,041	1,622	2,487

The sum of squares for these four groups may be divided into three parts explained by first, second, and third degree polynomial regressions. The coefficients and divisors are taken from the table ($k = 15$):

$$1. \frac{[-3(2,313) - 1(2,041) + 1(1,622) + 3(2,487)]^2}{(20)(15)} = 35$$

$$\begin{array}{rcl}
 2. \frac{[+1(2,313) - 1(2,041) - 1(1,622) + 1(2,487)]^2}{(4)(15)} & = & 21,546 \\
 3. \frac{[-1(2,313) + 3(2,041) - 3(1,622) + 1(2,487)]^2}{(20)(15)} & = & 6,826 \\
 \text{Total} & & 28,407
 \end{array}$$

It is plain that the linear trend is practically horizontal; that the deviations are fitted quite well by a parabola; that deviations from parabolic trend are significant, since $F = 6,826/54$, $d.f. = 1$ and 24 ; and that there is no deviation from the cubic. The third statement is based on either of two facts: (i) the total sum of squares is completely accounted for by the three regressions; or (ii) a cubic may be fitted to any 4 points.

EXAMPLE 15.15—Taking days 9–14, inclusive, from the chick embryo data of sections 14.2 and 14.6, divide the sum of squares of the six weights, 0.75541, into five portions, one corresponding to a polynomial of each degree through the fifth.

EXAMPLE 15.16—In the search for causes of lodging in oats, Brady (3) measured the thickness of the sclerenchyma cell wall in stems planted 4, 6, and 8 inches apart. Twenty-seven plots were planted at each spacing. The mean thicknesses of cell walls were 4.34, 4.69, and 4.67 units. In the analysis of variance, the mean squares for spacing and error led to $F = 1.0465/0.3482 = 3.01$, not quite reaching the 5% point for $d.f. = 2$ and 56 . (Three varieties each with three spacings were planted in a 9×9 latin square, hence the 56 degrees of freedom for error.) Show that the regression of thickness on spacing is significant, $F = 4.22$, $d.f. = 1$ and 56 . This is always an interesting possibility. Any one of a group of comparisons may attain significance even though the group as a whole falls short.

EXAMPLE 15.17—If you wish to duplicate the partitioning of the sum of squares of the chick embryo data by using the present method, you will need the coefficients for eleven items. From the references cited, I have copied down those for the first four polynomials, together with the embryo dry weights. The answers are at the foot of each column:

Dry Weight	Coefficients for Polynomial of Degree			
	1	2	3	4
0.029	-5	15	-30	6
0.052	-4	6	6	-6
0.079	-3	-1	22	-6
0.125	-2	-6	23	-1
0.181	-1	-9	14	4
0.261	0	-10	0	6
0.425	1	-9	-14	4
0.738	2	-6	-23	-1
1.130	3	-1	-22	-6
1.882	4	6	-6	-6
2.812	5	15	30	6
Divisor	110	858	4,290	286
Sum of squares	6.0785	1.8432	0.2368	0.0060

15.7—Applications to randomized blocks experiments. Among the many experiments conducted in randomized blocks, there are some for which the straightforward analysis of variance is not conclusive. One instance is the experiment contrasting control measures, the data being numbers of individuals occurring in the plots. Such data are often skewed instead of normal, or exhibit heterogeneous variance, or have means not independent of variance. This was discussed in section 11.17, and will be considered further in section 16.7, but one method of handling the situation may be given here.

Some data on weed infestation in cereals, reported by Bartlett (2) and examined in detail by Cochran (6), are rearranged by treatment in table 15.2. While treatment differences are obviously significant, the error variance is not suitable for making critical tests—it is too small for the highly variable treatments, *A* and *B*, and too large for the others. This may be inferred from the ranges in the last line of the table.

TABLE 15.2
NUMBER OF POPPY PLANTS IN OATS
Plants per $3\frac{3}{4}$ square feet.

Block	Treatment					
	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>
1	438	538	77	17	18	115
2	442	422	61	31	26	57
3	319	377	157	87	77	100
4	380	315	52	16	20	45
Sum	1,579	1,652	347	151	141	317
Range	123	223	105	71	59	70

Sums of squares: Treatment, 641,023; Error, 39,801. Mean square for error, 2,653.

The nature of the treatments is not reported, except that *A* was “no treatment.” The data clearly form three pairs, and we shall assume that this reflects some relationship in the character of the treatments. The method used is based on the fact that the testing of paired data is little affected by heterogeneity of variance.

The following comparisons are indicated in the data:

	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>
1. <i>A</i> vs. <i>B</i>	—	+	0	0	0	0
2. <i>C</i> vs. <i>F</i>	0	0	+	0	0	—
3. <i>D</i> vs. <i>E</i>	0	0	0	+	—	0
4. <i>C</i> + <i>F</i> vs. <i>D</i> + <i>E</i>	0	0	+	—	—	+
5. <i>A</i> + <i>B</i> vs. others	2	2	—	—	—	—

These five orthogonal comparisons are applied to the data in *each block* with the results given in table 15.3. As examples: (i) for comparison 1 in block 1, $-A + B = -438 + 538 = 100$; (ii) for comparison 4 in block

TABLE 15.3
DIFFERENCES FOR 5 ORTHOGONAL COMPARISONS ON EACH BLOCK, AND MEAN SQUARES
FOR TESTING SIGNIFICANCE

Block	Comparison				
	1	2	3	4	5
1	100	-38	-1	157	1,725
2	-20	4	5	61	1,553
3	58	57	10	93	971
4	-65	7	-4	61	1,257
Total	73	30	10	372	5,506
Divisor	2	2	2	4	12
Mean square:					
Difference	666	112	12	8,649	631,584
Error	2,776	756	20	512	9,203

4, $+C - D - E + F = 52 - 16 - 20 + 45 = 61$; (iii) for comparison 5 in block 1, $2A + 2B - C - D - E - F = 876 + 1,076 - 77 - 17 - 18 - 115 = 1,725$.

In each comparison the total is the same as if the difference had been calculated from the sums in table 15.2.

The mean square for the comparison difference is calculated from the total as usual. Thus, for comparison 4,

$$\frac{(372)^2}{(4)(4)} = 8,649$$

For error, the corrected sum of squares is calculated from the block differences, then divided by both the indicated divisor and the 3 degrees of freedom for 4 blocks: $k = 1$. The steps are, for comparison 1:

1. $Sx^2 = (100)^2 + \dots + (-65)^2 - (73)^2/4 = 16,657$
2. Sum of squares $= 16,657/2 = 8,328$
3. Mean square $= 8,328/3 = 2,776$

A final note on the computation: the five mean squares for treatment differences, each with 1 *d.f.*, add to 641,023, the same as the sum of squares in table 15.2; also, the error mean squares, each the result of dividing a sum of squares by 3 *d.f.*, add to 13,267, which is 1/3 of the original sum of squares for error as it should be. This verifies both the calculations and the orthogonality of the comparisons.

Anticipated differences among the error mean squares turn out to be highly significant (section 10.13), showing that the average error, 2,653, of table 15.2 was unsuited for any comparison except the first. Since it is assumed that comparison 4,

$$C + F - (D + E) = 372$$

reflects a feature of the design (not merely a grouping of fortuitous data), it is to be noted that this difference is highly significant when tested against its correct error, whereas it would have appeared nonsignificant if compared to the average error. Quoting Cochran: "This example illustrates the most important condition for the application of a combined analysis

TABLE 15.4
WEIGHT (OUNCES) OF ASPARAGUS CUT BEFORE JUNE 1 FROM PLOTS WITH
VARIOUS CUTTING TREATMENTS

Block	Year	Cutting Ceased				Total	Treatment Components		
		June 1	June 15	July 1	July 15		Linear T_L	Quadratic T_Q	Cubic T_C
1	1930	230	212	183	148	773	-275	-17	-5
	1931	324	415	320	246	1,305	-329	-165	-207
	1932	512	584	456	304	1,856	-752	-224	-176
	1933	399	386	255	144	1,184	-896	-98	-138
		1,465	1,597	1,214	842		-2,252	-504	-526
2	1930	216	190	186	126	718	-274	-34	+78
	1931	317	296	295	201	1,109	-349	-73	+113
	1932	448	471	387	289	1,595	-561	-121	-93
	1933	361	280	187	83	911	-927	-23	-1
		1,342	1,237	1,055	699		-2,111	-251	+97
3	1930	219	151	177	107	654	-310	-2	+190
	1931	357	278	298	192	1,125	-475	-27	+225
	1932	496	399	427	271	1,593	-647	-59	+309
	1933	344	254	239	90	927	-777	-59	+209
		1,416	1,082	1,141	660		-2,209	-147	+933
4	1930	200	150	209	168	727	-37	+9	+209
	1931	362	336	328	226	1,252	-416	-76	+112
	1932	540	485	462	312	1,799	-707	-95	+159
	1933	381	279	244	168	1,072	-674	+26	+108
		1,483	1,250	1,243	874		-1,834	-136	+588
Total		5,706	5,166	4,653	3,075	18,600	-8,406	-1,038	1,092

Preliminary Analysis of Variance

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Blocks	3	30,170	10,057
Years	3	518,722	172,907
BY	9	5,092	566
Treatments	3	241,377	80,459
TB	9	21,861	2,429
TY	9	51,178	5,686
TBY	27	10,395	385

of variance: *the treatment comparisons must be such that they may all be reasonably expected to have the same variance.*"

Other applications of individual comparisons are useful if treatments at several levels are tried in randomized blocks, or if trends in the yields of a perennial crop are the focus of interest. Both occur in an experiment by Haber (10) to compare the effects of various cutting treatments on asparagus. Planting was in 1927 and cutting began in 1929. One plot in each block was cut till June 1, others to June 15, July 1, and July 15. The yields in table 15.4 are the weights cut to June 1 in every plot, irrespective of later cuttings in some of them. This weight, then, is a measure of vigor, and the experiment is to determine the trend through 4 years.

The total yields throughout the 4 years, tabulated in a 4×4 treatment-block table, give lines 1, 4, and 5 in the analysis of variance. As in section 11.16, the treatment-block interaction is the appropriate error for testing treatments, so that the means and mean squares for this 4×4 table might be the end-point of the investigation.

But we wish to examine the trends in yield, not only those caused by increasing severity of cutting, but those that may be associated with loss of vigor through the years. To this end, the linear regressions of yield on treatment are calculated, not only for the totals, but also for every block and year separately. On the assumption of equal intervals between cutting dates, the coefficients for 4 groups (table 15.1) are available. Thus, the linear component for block 1 in 1930 is

$$-3(230) - (212) + (183) + 3(148) = -275$$

That is, $S_{xy} = -275$ if x is taken as $-3, -1, 1$, and 3 . The regression coefficient is $b = S_{xy}/S_x^2 = -275/20 = -13.75$ ounces per unit of x . (Note: there are 2 units of x for each treatment interval.)

The regression of yield is uniformly downward (negative). Moreover, examination of the linear components indicates that this trend becomes more pronounced as the age of the plants increases. Perhaps these features will be more evident if the linear components, T_L , are tabulated in the 4×4 treatment-block table 15.5. The totals in the last line emphasize the steepening downward slope as the years pass: the decrease in vigor with increasing severity of cutting is accentuated as the plants grow older.

Keep in mind that all the entries in table 15.5 are sums of products, S_{xy} . The total, $-8,406$, could be got directly from the treatment yields (table 15.4):

$$-3(5,706) - (5,166) + (4,653) + 3(3,075) = -8,406$$

The average, $(-8,406)/16 = -525.4$, leads to the regression of yield on treatment, $b = S_{xy}/S_x^2 = -525.4/20 = -26.27$ ounces per unit $x = -52.54$ ounces per cutting interval (15 days). The associated reduction in sum of squares is indicated in the table. The corresponding sum of squares for blocks, calculated as explained earlier in this section, results

TABLE 15.5
 LINEAR COMPONENTS, T_L , OF REGRESSIONS, YIELD ON DATE OF CUTTING

Block	Year of Harvest				Total
	1930	1931	1932	1933	
1	-275	-329	-752	-896	-2,252
2	-274	-349	-561	-927	-2,111
3	-310	-475	-647	-777	-2,209
4	-37	-416	-707	-674	-1,834
Total	-896	-1,569	-2,667	-3,274	-8,406

$$T_L: \frac{(-8,406)^2}{(20)(16)} = 220,815$$

$$T_{LB}: \frac{(-2,252)^2 + \dots + (-1,834)^2}{(20)(4)} - T_L = 1,323$$

$$T_{LY}: \frac{(-896)^2 + \dots + (-3,274)^2}{(20)(4)} - T_L = 42,892$$

$$T_{LBY}: \left(\frac{(-275)^2 + \dots + (-674)^2}{20} - T_L \right) - T_{LB} - T_{LY} = 4,756$$

Analysis of Variance of Linear Components of Treatments, T_L

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
T_L	1	220,815	220,815
T_{LB} (Error)	3	1,323	441
Total	4	222,138	

in the error mean square, 441, against which T_L may be tested. Two points are worthy of notice: most of the treatment sum of squares, 241,377 (table 15.4) is due to linear regression; and the mean square, $T_{LB} = 441$, is greatly reduced from $TB = 2,429$. The small value of T_{LB} leads one to observe the uniformity of the block totals in table 15.5 and to look for the larger sources of variation in TB .

Notes on computation: (i) The correction term for T_{LB} turns out to be T_L . (ii) Since the coefficients, $-3, -1, +1, +3$, are used to get all the entries in table 15.5, the sum of their squares, 20, appears as divisor in each fraction. (iii) The other factor in the denominator is, as usual in such a table, the number of items in the sum which is squared in the numerator.

The treatment-year interaction in the preliminary analysis of variance is highly significant, $F = 5,686/385 = 14.77$. The linear component of treatment accounts for a large part of this interaction, as shown by com-

parison of $T_L Y = 42,892$ with $TY = 51,178$. We wish to segregate a portion of TBY suitable for testing $T_L Y$, and to learn if the remainder of TY is homogeneous.

Pursuing the objectives mentioned in the first and third paragraphs preceding, table 15.6 is constructed by applying the coefficients for 4

TABLE 15.6
REGRESSION OF LINEAR COMPONENTS ON YEARS, $T_L Y$

Block	Linear, $T_L Y_L$	Quadratic, $T_L Y_Q$	Cubic, $T_L Y_C$
1	-2,286	- 90	+648
2	-2,171	-291	- 17
3	-1,573	+ 35	+ 49
4	-2,202	+412	+236
Total	-8,232	+ 66	+916

groups (table 15.1) to the linear components in table 15.5. As examples in block 1:

$$T_L Y_L = -3(-275) - (-329) + (-752) + 3(-896) = -2,286$$

$$T_L Y_Q = (-275) - (-329) - (-752) + (-896) = - 90$$

$$T_L Y_C = - (-275) + 3(-329) - 3(-752) + (-896) = 648$$

The entries in this table indicate the regressions on years (the trends) of the linear components of treatment regressions. Thus, the average, $-8,232/4 = -2,058$, gives the linear regression coefficient, $-2,058/20 = -102.9$ ounces per half year. The sum of squares for this regression is

$$T_L Y_L: \frac{(-8,232)^2}{(20)(20)(4)} = 42,354$$

The denominator contains a factor, 20, for each application of the coefficients and 4 because the numerator is the sum of 4 values of $T_L Y_L$. Since this sum of squares contains most of $T_L Y = 42,892$, it is unnecessary to look further for sources of variation. Nevertheless, to illustrate the method, we calculate the sum of squares of the deviations from linear regression,

$$T_L Y_Q: \frac{(66)^2}{(20)(4)(4)} = 14$$

The divisors, 20 and 4, are the same as those in $T_L Y_L$, but the sum of the squares of the quadratic coefficients is 4 instead of 20. Finally, the sum of squares of the deviations from second degree regression is the *cubic component*,

$$T_L Y_C: \frac{(916)^2}{(20)(20)(4)} = 524$$

TABLE 15.7

ANALYSIS OF VARIANCE OF COMPONENTS OF THE TREATMENT EFFECTS—ASPARAGUS DATA

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Linear Components, T_L			
T_L	1	220,815	220,815
$T_L B$	3	1,323	441
$T_L Y_L$	1	42,354	42,354
$T_L Y_Q$	1	14	14
$T_L Y_C$	1	524	524
$T_L B Y_L$	3	802	267
$T_L B Y_Q$	3	3,283	1,094
$T_L B Y_C$	3	671	224
Sum	16	269,786	
Quadratic Components, T_Q			
T_Q	1	16,835	16,835
$T_Q B$	3	5,485	1,828
$T_Q Y_L$	1	744	744
$T_Q Y_Q$	1	6,440	6,440
$T_Q Y_C$	1	414	414
$T_Q B Y_L$	3	927	309
$T_Q B Y_Q$	3	2,117	706
$T_Q B Y_C$	3	89	30
Sum	16	33,051	
Cubic Components, T_C			
T_C	1	3,726	3,726
$T_C B$	3	15,053	5,018
$T_C Y_L$	1	536	536
$T_C Y_Q$	1	135	135
$T_C Y_C$	1	16	16
$T_C B Y_L$	3	507	169
$T_C B Y_Q$	3	880	293
$T_C B Y_C$	3	1,120	373
Sum	16	21,973	
Total	48	324,810	

One of the divisors, 20, is the same as before, but the other is the sum of the squares of the coefficients of the third degree regression, $(-1)^2 + (3)^2 + (-3)^2 + (1) = 20$. (If there were 5 years instead of 4, this would be 10; see table 15.1). We now verify the orthogonality of $T_L Y_L$, $T_L Y_Q$, and $T_L Y_C$ by addition of the sums of squares:

$$42,354 + 14 + 524 = 42,892$$

For testing the three mean squares just isolated, there are corresponding components of T_LBY . These are calculated from the columns of table 15.6 in the usual manner; for instance, in the middle column,

$$T_LBY_Q: \frac{(-90)^2 + \dots + (412)^2}{(20)(4)} - T_LY_Q = 3,283$$

The sum of this and the other two components, 802 and 671, is $T_LBY = 4,756$.

Summarized in the first part of table 15.7 are the computations described above. Among the regressions of T_L on years, only the linear component is significant. Since the 3 components of T_LBY do not differ significantly (Bartlett's test, section 10.13, gives $\chi^2 = 2.12$, $d.f. = 2$), it is right to use the mean square, $4,756/9 = 528$, to test all the components of T_LY .

Some heterogeneity remains to be located, so we proceed to isolate the quadratic and cubic components of T . The procedure follows closely that already described, only the coefficients and corresponding divisors being changed. For

$$T_Q: \frac{(-1,038)^2}{(4)(16)} = 16,835$$

This is the reduction in sum of squares due to fitting a second degree polynomial (parabola) to the deviations from linear regression. For

$$T_C: \frac{(1,092)^2}{(20)(16)} = 3,726$$

Verification: $T_L + T_Q + T_C = 220,815 + 16,835 + 3,726 = 241,376 = T$.

The corresponding components of TB , T_QY , and T_QBY are calculated from a table like 15.5 except that its entries are the quadratic components, T_Q . A third table with T_C yields T_CB , T_CY , and T_CBY . From each of these tables is made one like 15.6 and from them the computations are completed (table 15.7). The results are all recorded for the guidance of the student though only a few of them indicate more than sampling variation. We shall look at these few and condense the table into a form which would be suitable for publication—table 15.8.

First, look at the components of TB (mean squares):

$$\begin{aligned} T_LB &= 441 \\ T_QB &= 1,828 \\ T_CB &= 5,018 \end{aligned}$$

Despite the discrepancies, the components do not differ significantly, chi-square being at about the 20% point. If you plot the four treatment means in each block, you will observe the following: (i) The linear regressions of yield on treatment are quite uniform, especially in the first three blocks (observe the sums of products in column T_L of table 15.4, $-2,252$,

TABLE 15.8
FINAL ANALYSIS OF VARIANCE OF THE ASPARAGUS DATA

Source of Variation	Degrees of Freedom	Mean Square
Blocks	3	10,057
Years	3	172,907
<i>BY</i>	9	566
Treatments:		
Linear, T_L	1	220,815
Quadratic, T_Q	1	16,835
Error for treatments, TB	9	2,429
Trends:		
Linear, $T_L Y_L$	1	42,354
Quadratic, $T_Q Y_Q$	1	6,440
Error for trends, TBY	27	385
Homogeneous effects	8	
Total	63	

etc.), hence the small $T_L B$. (ii) The parabolic deviations are noticeable in block 1, but decrease in the other blocks. Treatment 1 is uniformly high in all blocks and treatment 4 is low, but the relative positions of the intermediate treatments give rise to the variations in $T_Q B$ and $T_C B$. Cutting until June 15 did not affect the vigor of the plants in block 1, but in block 3 this treatment was more severe than the cutting until July 1. Apparently, all this can be attributed to the available nutrients in the plots. (iii) The notable differences in the cubic components, from -526 in block 1 to $+933$ in block 3, are due chiefly to the behavior of plots 2 and 3. Since there is a tendency toward uniformity from year to year within the blocks, this also seems traceable to differential plot fertility.

The data suggest that the relation between vigor and date of final cutting is critically associated with fertility. But fertility was not controlled in this experiment; hence, the appropriate error for treatments should include all components of TB . This reasoning, added to the non-significance of the differences among the components, leads us to use the pooled mean square, 2,429, as error for all treatment components.

As for the nine components of the second order interaction, they are notably homogeneous, chi-square being near the 40% point. The error for all first order interactions, then, is taken as $TBY = 385$.

The obvious linear decrease in vigor with prolongation of date of final cutting, T_L , and its equally obvious downward trend, $T_L Y_L$, have been commented on. There remain only two effects to be discussed.

1. The deviations of the treatment means from their linear regression show the increasingly detrimental effects of prolonged cutting. The means, from the row of totals in table 15.4, are 356.6, 322.9, 290.8, and 192.2. The successive differences measure the departure from linearity; 33.7, 32.1, and 98.6. The augmented decrease in yield for the July 15

cutting is a feature of all blocks except the first. $T_Q = 16,835$, a significant variance, is the average square of that portion of the deviations from linearity explained by parabolic regression. A graph of the treatment means makes this feature more vivid.

2. The highly significant $T_Q Y_Q = 6,440$ results from a parabolic trend in the quadratic components of the treatment regressions. To see this, accumulate T_Q by years:

1930	1931	1932	1933
-44	-341	-499	-154

These quadratic portions of the deviations of the treatment yields from linear regression, T_Q , do not exhibit much linear trend, but do show pronounced deviations which are well described as parabolic. Straight lines fitted the treatment yields fairly well in 1930 and 1933, but in the two intermediate years the June 15 and July 1 dates had yields well above the linear regressions. The interpretation is that in the more favorable seasons of 1931 and 1932 the effects of the June 15 and July 1 cuttings were relatively less harmful than in the years of low yield, 1930 (the plants were young, too) and 1933.

EXAMPLE 15.18—Notice the total sum of squares for 48 degrees of freedom in table 15.7. Of what parts of the analysis of variance in table 15.4 is this a re-analysis?

EXAMPLE 15.19—Verify and fix in mind the meaning of an equation like the following, the terms being sums of squares:

$$T_Q Y_L + T_Q Y_Q + T_Q Y_C = T_Q Y$$

EXAMPLE 15.20—For each year, calculate the regression of yield on treatment date, using as a unit 15 days: 1 corresponds to June 1, 2 to June 15, etc. For 1930, $\bar{Y} = 179.5 - 22.40(X - 2.5)$. Plot each regression together with the treatment means. Verify the explanation of $T_Q Y_Q$.

EXAMPLE 15.21—To the treatment yields for 1932 fit linear, quadratic and cubic regressions as in section 14.6.

15.8—Switch-back or reversal experiments. Brandt (4) has shown that individual comparisons are convenient for evaluating the results of those experiments where two treatments are tried alternately on two groups of animals. Such are especially common in investigations of the effect of treatment on milk flow. An example will serve to explain the situation.

The milk yields in table 15.9 are data from an experiment (1) to determine the value of a gluten supplement to the grain mixture in feeding dairy cows. In order to obviate the effects of the normal decline in milk flow, the feeds were tried alternately in three periods of 35 days each. As indicated in the table, lot *A* received the grain and gluten (*GG*) during the first period, while lot *B* had only grain (*G*). In the second period these parts of the rations were interchanged.

The comparisons are by period. Hence, for each animal,

$$\begin{array}{rcl} a & 0 & -c \\ \text{and} & a & -2b + c \end{array}$$

TABLE 15.9
POUNDS OF MILK PRODUCED BY TWO LOTS OF HOLSTEIN-FRIESIANS DURING EACH OF THREE PERIODS. TWO RATIONS, ONE CONTAINING GRAIN (*G*), THE OTHER GRAIN AND A GLUTEN FEED (*GG*)

Lot	Cow Number	Period			Comparison $a - 2b + c$	Sum
		<i>a</i>	<i>b</i>	<i>c</i>		
<i>A</i>	1	(<i>GG</i>) 671	(<i>G</i>) 610	(<i>GG</i>) 597	48	215
	2	615	555	488	- 7	
	3	777	733	694	5	
	4	1,102	959	940	124	
	5	764	718	717	45	
<i>B</i>	6	(<i>G</i>) 433	(<i>GG</i>) 414	(<i>G</i>) 363	-32	-219
	7	745	797	780	-69	
	8	858	754	680	30	
	9	977	1,025	1,007	-66	
	10	655	616	495	-82	

give complete information. The first comparison measures the obvious decrease in milk flow, not part of the object of this experiment. The second, however, gives the deviation of the milk yield in the middle period as measured from the linear regression. Its value for each cow is recorded in the table, sums for each lot appearing in the last column. For lot *A* these sums are (*GG*) - (*G*), while for lot *B* they are the negatives, (*G*) - (*GG*). The difference, therefore,

$$215 - (-219) = 434,$$

represents the total pounds of milk attributable to the gluten feed, assuming no carry-over from the preceding period. Brandt showed that this difference is properly tested for significance by applying the methods of sections 4.2 and 15.2. The error variance comes from the pooling of the sums of squares in the two lots of comparisons. In lot *A*, for example,

$$Sx^2 = (48)^2 + (-7)^2 + (5)^2 + (124)^2 + (45)^2 - (215)^2/5 = 10,534$$

The corresponding sum for lot *B* is 8,173. The error variance is, then,

$$\frac{10,534 + 8,173}{4 + 4} = 2,338$$

For the comparison of gluten and nongluten supplements, the sum of squares is,

$$\frac{[215 - (-219)]^2}{2(5)} = 18,836$$

Hence, $F = 18,836/2,338 = 8.06$, significant for *d.f.* = 1 and 8.

This ingenious method of testing significance in reversal trials may be

extended to any number of trial periods. If there were four periods, for example, the comparison for third degree polynomial, table 15.1, would be used.

EXAMPLE 15.22—In table 15.9, run up the sums in the six groups of milk yields and represent them graphically. In lot *A*, the sum in period *b* falls below the line connecting the sums in *a* and *c*. The opposite condition is found in lot *B*.

EXAMPLE 15.23—Cannon, Hansen, and O'Neil (5) tried the effect which two methods of watering dairy cows, inside the barn (*I*) and outside (*O*), might have on butterfat production. The pounds of butterfat yielded by ten cows in two lots at each of four periods are tabulated:

Lot	Period			
	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>
<i>A</i>	(<i>O</i>)	(<i>O</i>)	(<i>I</i>)	(<i>O</i>)
	40	32	28	14
	27	19	20	14
	35	27	31	25
	30	28	27	26
	37	30	32	27
<i>B</i>	(<i>O</i>)	(<i>I</i>)	(<i>O</i>)	(<i>I</i>)
	51	55	40	40
	32	32	32	28
	36	38	29	31
	32	30	25	25
	21	22	15	15

Compute for each cow the comparison for four periods (table 15.1),

$$-a + 3b - 3c + d$$

The sums are for lot *A*, -69, and for lot *B*, 75. Calculate $F = 2,074/132.6 = 15.6$, highly significant for *d.f.* = 1 and 8.

15.9—Factorial experiments. A comprehensive and efficient kind of experiment, one that will probably increase in popularity, is the *factorial* (7) (12), in which two or more series of treatments are tried in all combinations. For example, three sources of protein might be used in a feeding experiment, each at two levels. The six combinations of this pair of series, source (*s*) and level (*l*), are

$$\begin{array}{ccc} s_1l_1 & s_2l_1 & s_3l_1 \\ s_1l_2 & s_2l_2 & s_3l_2 \end{array}$$

Each combination might be applied to a lot of, say, 10 animals, with the resulting analysis of degrees of freedom,

$$\begin{array}{rr} \text{Treatment combinations,} & 5 \\ \text{Error} & 54 \end{array}$$

Information about the effect of level is furnished by every animal, because 30 receive the low level and 30 the high. The comparison of levels corre-

sponds with one of the five degrees of freedom for treatments. Furthermore, the source of protein might be selected for two orthogonal comparisons, such as

$$\begin{aligned} -s_1 + s_3 \\ s_1 + s_3 - 2s_2 \end{aligned}$$

The remaining two degrees of freedom correspond to the interactions in the 2×3 table.

One form of the factorial experiment is that illustrated in table 15.10 in which each series of treatments is at two levels; in this case, none and

TABLE 15.10

YIELDS OF WHEAT IN A 2^3 FACTORIAL EXPERIMENT LAID DOWN IN A LATIN SQUARE. NITROGEN (*n*), PHOSPHOROUS (*p*), AND POTASSIUM (*k*) WERE TRIED AT TWO LEVELS, NONE AND SOME. (1) INDICATES NO FERTILIZER

<i>p</i>	<i>n</i>	<i>np</i>	<i>k</i>	<i>nk</i>	(1)	<i>npk</i>	<i>pk</i>
18.8	12.2	18.3	15.8	11.4	11.5	19.4	18.9
<i>n</i>	<i>nk</i>	<i>pk</i>	<i>npk</i>	<i>p</i>	<i>k</i>	<i>np</i>	(1)
12.9	7.3	17.4	17.2	19.7	12.0	19.0	15.6
<i>nk</i>	<i>np</i>	<i>n</i>	<i>p</i>	(1)	<i>npk</i>	<i>pk</i>	<i>k</i>
10.7	17.5	10.4	18.0	9.8	16.6	17.5	14.3
<i>pk</i>	<i>k</i>	<i>npk</i>	(1)	<i>n</i>	<i>np</i>	<i>p</i>	<i>nk</i>
18.3	12.6	14.2	12.2	11.4	14.5	16.9	16.1
<i>np</i>	(1)	<i>nk</i>	<i>n</i>	<i>pk</i>	<i>p</i>	<i>k</i>	<i>npk</i>
17.9	12.8	13.3	11.3	16.5	15.6	10.9	16.7
<i>k</i>	<i>pk</i>	(1)	<i>np</i>	<i>npk</i>	<i>n</i>	<i>nk</i>	<i>p</i>
14.9	18.2	12.8	17.1	15.8	9.5	8.9	20.6
<i>npk</i>	<i>p</i>	<i>k</i>	<i>pk</i>	<i>np</i>	<i>nk</i>	(1)	<i>n</i>
19.0	18.9	11.2	17.1	17.9	8.6	10.2	14.5
(1)	<i>npk</i>	<i>p</i>	<i>nk</i>	<i>k</i>	<i>pk</i>	<i>n</i>	<i>np</i>
17.5	20.4	20.8	16.4	16.8	18.5	13.6	23.0
Treatment Sums							
(1)	102.4	<i>p</i>	149.3	<i>np</i>	145.2	<i>pk</i>	142.4
<i>n</i>	95.8	<i>k</i>	108.5	<i>nk</i>	92.7	<i>npk</i>	139.3
Preliminary Analysis of Variance							
Source of Variation	Degrees of Freedom		Sum of Squares		Mean Square		
Treatments	7		513.79		73.4		
Error	42				2.190		

some. The three series—nitrogen, phosphorous, and potassium fertilizers—make $2^3 = 8$ combinations,

$n_0p_0k_0$	$n_0p_0k_1$
$n_0p_1k_0$	$n_0p_1k_1$
$n_1p_0k_0$	$n_1p_0k_1$
$n_1p_1k_0$	$n_1p_1k_1$

The symbols for the combinations are usually abbreviated, as shown in the table, the check (control or no treatment) being represented by (1).

Factorial experiments may be laid out in any convenient design, such as randomized blocks; in the one illustrated, Goulden (9) used the 8×8 latin square. The experimental error, 2.190, $d.f. = 42$, is available for testing each of the 7 treatment comparisons. Sometimes there is no replication, the assumption being that one or more of the higher order interactions contains only the random variation suitable for estimating error (see section 11.14).

The treatment sums, shown in the table, may be compared as follows:

Effects of phosphorous:

acting alone	$p - (1) = 149.3 - 102.4 = 46.9$
with nitrogen,	$pn - n = 145.2 - 95.8 = 49.4$
with potash,	$pk - k = 142.4 - 108.5 = 33.9$
with both,	$pnk - nk = 139.3 - 92.7 = 46.6$
Total	176.8

It is evident that phosphorous always increased the yield of wheat on these plots. Since there were 32 plots at each level, the average increase per pair was $176.8/32 = 5.525$. This mean difference has the standard error,

$$s_{\bar{d}} = \sqrt{2(2.190)/32} = 0.3700,$$

so that $t = 5.525/0.3700 = 14.93$ leaves no doubt about the effectiveness of phosphorous under the conditions of this experiment.

But, you may object, the phosphorous comparison has been made in four different circumstances so that the average effect may not be fully informative. The objection is valid if the mean differences (46.9/8, etc.), 5.86, 6.18, 4.24, and 5.82, reflect the presence of the other fertilizers. As a preliminary test, compare the range, $6.18 - 4.24 = 1.94$, with its 5% level, table 5.5. Since $s_{\bar{d}} = \sqrt{2(2.190)/8} = 0.740$, the 5% level of range is approximated by $(0.740)(3.56) = 2.70$; hence, the mean differences may be random samples from a common population. Some more discriminating tests will now be discussed.

Begin by looking at the difference between the effects of phosphorous (i) acting alone and (ii) in the presence of nitrogen, $49.4 - 46.9 = 2.5$. We have noticed this kind of difference before, and have called it *interaction*. Continuing, we have,

Interactions of phosphorous:

with nitrogen alone	$49.4 - 46.9 = 2.5$
with nitrogen in the presence of potash	$46.6 - 33.9 = 12.7$
Total	<u>15.2</u>

Assuming again that these interactions differ only as in sampling, the average interaction is $15.2/32 = 0.475$, so that $t = 0.475/0.3700 = 1.28$, nonsignificant.

But perhaps the two interactions differ significantly. As we have had occasion to note before (section 11.14), the difference between two interactions is called *second order interaction*, the interaction among the three factors, phosphorous, nitrogen and potash. The average difference is only $(12.7 - 2.5)/32 = 0.319$; hence, $t = 0.319/0.370 = 0.86$ falls short of significance.

We have now learned that phosphorous uniformly increases yield under the circumstances of this experiment, and that its action may be unaffected by the addition of nitrogen or potash.

The effects of nitrogen may be examined in the same way, but a more compact scheme (table 15.11) is adequate after you have become accustomed to the reasoning involved. Each difference, recorded under N , is the effect of nitrogen alone or in the presence of p , k , or pk . Since the

TABLE 15.11
EFFECTS OF NITROGEN IN WHEAT EXPERIMENT

Treatment	Sum of Yields	N	NP	NPK
(1)	102.4	— 6.6	2.5	10.2
n	95.8			
p	149.3	— 4.1		
np	145.2			
k	108.5	—15.8	12.7	
nk	92.7			
pk	142.4	— 3.1		
npk	139.3			
Total		—29.6	15.2	10.2

differences do not vary significantly, they are combined in the lower line of the table. The totals represent the effects of nitrogen, N ; the interaction, NP ; and the second order interaction NPK : the latter two are identical with the interactions in the phosphorous series. Notice that the lower case letters represent the treatments, while the capitals indicate effects or differences.

Instead of dividing each sum by 32 and computing t , it is easier to calculate the significant difference between sums, using $t_{.05} = 2.018$:

$$2.018\sqrt{(32)(2)(2.190)} = 23.89$$

Comparison of this with N shows that the nitrogen effect is a bit larger than the 5% level of significance: the addition of nitrogen may be expected to depress yields in the population of which this sample is representative.

The test of NPK is the test of the significance of the difference between the two components of NP , the interaction of nitrogen and phosphorous (i) alone, and (ii) in the presence of potash. If NPK is evidently a population characteristic the investigator is warned that the nitrogen-phosphorous interaction is different with potash and without, so that the average effect, NP , may not be informative. He may make one recommendation about n and p if k is to be used and a different one if k is not; or he may have to warn the farmer, for example, that n and p should be used together only if k is added.

If NPK can be assumed negligible in the population; that is, if the two parts of NP are regarded as estimates of the same population difference, then NP is a homogeneous estimate of the difference between the effects of nitrogen, (i) with phosphorous and (ii) without, irrespective of the presence or absence of potash. In making this statement, the effect of potash is ignored—it may be great or it may be, as in the present experiment, negligible. The statement implies that, for evaluating the interaction, NP , the upper and lower halves of the table may be superimposed. If NP is negligible, as here, the recommendation about nitrogen can be made without reference to phosphorous; but if NP were large, recommendations about the two fertilizers would have to be related. In other words, in the first case the average effect of N is the appropriate estimate; whereas, in the second case the average effect may not be interesting—the effect with phosphorous would be different from that without.

Three tables like 15.11 are necessary to isolate the seven orthogonal comparisons among the eight treatments. The results are summarized in table 15.12. Aside from N and P , none of the effects are significant. Note: three tables like 15.11 will give all the effects if the order of the treatments is properly selected. After (1)- n - p try (1)- p - k , then (1)- k - n .

Referring again to table 15.11, it will be evident that N is the differ-

TABLE 15.12
SEVEN EFFECTS AMONG THE EIGHT TREATMENT COMBINATIONS
IN THE WHEAT EXPERIMENT

Effect	Difference	Effect	Difference
N	-29.6	NK	- 8.2
P	176.8	PK	-15.8
K	- 9.8	NPK	10.2
NP	15.2		

5% significant difference, 23.89

ence between the yields of the 32 plots receiving nitrogen and the 32 without it:

$$(95.8+145.2+92.7+139.3) - (102.4+149.3+108.5+142.4) = -29.6$$

Again, *NP* is the difference between the yields of the 32 plots having both nitrogen and phosphorous or neither and of the 32 having either nitrogen or phosphorous but not both:

$$(102.4+145.2+108.5+139.3) - (95.8+149.3+92.7+142.4) = 15.2$$

Finally, *NPK* is the difference between the yields of the 32 plots having either one or three of the fertilizers applied as against the 32 plots having either none or two:

$$(95.8+149.3+108.5+139.3) - (102.4+145.2+92.7+142.4) = 10.2$$

From such rules, all the effects may be calculated without tables such as 15.11, but I do not advise it except as verification—some of the important features of the experiment may be overlooked.

The foregoing formulas are summarized in table 15.13. For the *main effects*, *N*, *P*, and *K*, mark with + all combinations having the fertilizer and with - those lacking it. The signs for the interactions are got by applying the rule for multiplication, like signs give +, unlike -. Thus:

$$\begin{array}{r} N: \quad - + - - + + - + \\ K: \quad - - - + - + + + \\ \hline NK: \quad + - + - - + - + \end{array}$$

Notice the rhythm in each line: it will help prevent mistakes. Verify the signs in the interactions by comparing them with the rules preceding.

The difference for each effect leads to a mean square in the manner

TABLE 15.13
COEFFICIENTS OF INDIVIDUAL COMPARISONS SUMMARIZING FORMULAS
FOR THE TREATMENT EFFECTS

Treatments	(1)	<i>n</i>	<i>p</i>	<i>k</i>	<i>np</i>	<i>nk</i>	<i>pk</i>	<i>npk</i>	Differ- ences
Sum of Yields	102.4	95.8	149.3	108.5	145.2	92.7	142.4	139.3	
Comparisons:									
<i>N</i>	-	+	-	-	+	+	-	+	-29.6
<i>P</i>	-	-	+	-	+	-	+	+	176.8
<i>K</i>	-	-	-	+	-	+	+	+	-9.8
<i>NP</i>	+	-	-	+	+	-	-	+	15.2
<i>NK</i>	+	-	+	-	-	+	-	+	-8.2
<i>PK</i>	+	+	-	-	-	-	+	+	-15.8
<i>NPK</i>	-	+	+	+	-	-	-	+	10.2

already familiar: $k = 8$ and the sum of the squares of the coefficients is also 8. Taking NP as an example,

$$\frac{(15.2)^2}{(8)(8)} = 3.61$$

This mean square is compared with experimental error by means of F :

$$F = 3.61/2.190 = 1.65$$

The test is, of course, identical with that using t . Verification: $\sqrt{1.65} = 1.28 = t$. In this factorial experiment the t test is much easier, especially if one uses the significant difference between sums. But as a check on the computations, as well as a matter of interest, the sum of the seven mean squares may be calculated and compared with 513.79 in table 15.10:

$$\frac{(-29.6)^2 + \dots + (10.2)^2}{(8)(8)} = 513.79$$

With experience and skill in the interpretation of data from factorials, construction of tables like 15.11 and 15.13 may be unnecessary. Yates (12) devised the ingenious method shown in table 15.14. The treatment sums

TABLE 15.14
YATES' DEVICE FOR COMPUTING EFFECTS

Treatments	Sums	1	2	3	Effects
(1)	102.4	198.2	492.7	975.6	
n	95.8	294.5	482.9	-29.6	N
p	149.3	201.2	-10.7	176.8	P
np	145.2	281.7	-18.9	15.2	NP
k	108.5	-6.6	96.3	-9.8	K
nk	92.7	-4.1	80.5	-8.2	NK
pk	142.4	-15.8	2.5	-15.8	PK
npk	139.3	-3.1	12.7	10.2	NPK
Total	975.6				

are entered in a characteristic order, any two treatments like n and p being followed by np , while k is then followed by nk and pk . In column 1 are successively:

4 sums like $95.8 + 102.4 = 198.2$

4 differences like $95.8 - 102.4 = -6.6$

Column 2 is made from column 1 in the same fashion:

4 sums like $294.5 + 198.2 = 492.7$

4 differences like $294.5 - 198.2 = 96.3$

Finally, column 3 is got from column 2 by the same sequence.

The first entry in column 3 is the total yield for the experiment, furnishing a check on the calculations to that point. Other methods of

verification may be found in Yates' article. The 7 differences may be tested by t or F as preferred.

The conclusion from this experiment is that in the region of which it is representative phosphorous alone may be recommended to increase the yield of wheat. Warning: do not proceed immediately to summarize a factorial experiment in a table like 15.13 or 15.14. First examine every detail of the comparisons in order to gain all the information.

EXAMPLE 15.24—The 2×2 table is often effective in studying the results of factorial experiments. The yields from a lime-inoculation trial on legumes might look like this:

	No Lime	Lime
No inoculation	40	60
Inoculated	80	200

If the entries are the sums from five replications, show that the average interaction is 10.

EXAMPLE 15.25—The results of a trial of nitrogen and phosphorous in ten replications might be like this:

	(1)	p
(1)	600	800
n	800	900

Show that average $NP = -5$, despite the greater yields of the plots having both.

REFERENCES

1. C. T. BAKER. Thesis submitted for the degree, Master of Science, Iowa State College (1927).
2. M. S. BARTLETT. Supplement to the Journal of the Royal Statistical Society, 3:68 (1936).
3. J. BRADY. Supplement to the Journal of the Royal Statistical Society, 2:99 (1935).
4. A. E. BRANDT. Iowa Agricultural Experiment Station Research Bulletin 234 (1938).
5. C. Y. CANNON, E. N. HANSEN, and JAMES R. O'NEAL. Iowa Agricultural Experiment Station Bulletin 292 (1932).
6. W. G. COCHRAN. The Empire Journal of Agriculture, 6:157 (1938).
7. R. A. FISHER. Design of Experiments. Oliver and Boyd, Edinburgh (1935-1942).
8. R. A. FISHER and F. YATES. Statistical tables. Oliver and Boyd, Ltd., Edinburgh (1938-1943).
9. C. H. GOULDEN. Methods of Statistical Analysis. John Wiley and Sons, Inc., New York (1939).
10. E. S. HABER. Journal of Agricultural Research, 45:101 (1932).
11. E. E. HOUSEMAN and R. D. ANDERSON. Iowa Agricultural Experiment Station Research Bulletin 297 (1942).
12. F. YATES. The design and analysis of factorial experiments. Imperial Bureau of Soil Science, Harpenden (1937).

Chapter 16

BINOMIAL AND POISSON DISTRIBUTIONS

16.1—Introduction. Small samples from these distributions were dealt with in chapters 1 and 9. We are now ready for the discussion of some features of large samples of enumeration data similar to the large sample methods for measurement data presented in chapter 8. Since in these distributions the mean and the variance are not independent, transformations may be necessary before analysis of variance is applied. Three of these are described in the closing section.

16.2—The binomial distribution. If the incidence of some event like survival or male sex is observed in *subsamples* of uniform size, the subsamples may be classified according to the number of times the event occurs. The resulting frequency distribution may be a sample from a *binomial distribution*, a distribution in which the probability of the occurrence of the event is the same for every individual.

As an example, suppose lots (subsamples) consisting of four full sister chicks are inoculated with an organism producing equal chances of life or death for every chick. The event of survival in the lots may occur any number of times from none to four. If 96 lots are injected, some such record as this may result:

Number of Surviving Chicks in Subsample	Number of Subsamples With Each Record of Survivals
0	1
1	20
2	41
3	25
4	9
Total	96

Clearly, no survivals in a lot, as well as four survivals, is unlikely. The modal number of living chicks is two.

The two specifications, four individuals in each subsample and 50% probability of survival, completely determine a binomial distribution. Most conveniently, the *relative frequency* of occurrence of 0, 1, 2, 3, and 4 survivors is given by the successive terms of the expansion of the binomial,*

* Look up the topics, *binomial theorem, probability, sampling, permutations and combinations* in any college algebra.

$$\begin{aligned}
 & \left(\frac{1}{2} + \frac{1}{2}\right)^4 \\
 &= \left(\frac{1}{2}\right)^4 + 4\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right) + 6\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2 + 4\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^3 + \left(\frac{1}{2}\right)^4 \\
 &= \frac{1}{16} + \frac{1}{4} + \frac{3}{8} + \frac{1}{4} + \frac{1}{16}
 \end{aligned}$$

This means that one-sixteenth of the subsamples in the population will have no survivals, one quarter will have one, etc. Combining this information with the sample fact of 96 lots, the *expected frequencies* in the classes are

$$(96)\left(\frac{1}{16}\right) = 6, \text{ etc.}; \text{ that is, } 6, 24, 36, 24, \text{ and } 6. \text{ The discrepancies be-}$$

tween these and the sample frequencies may be due to the vagaries of sampling or to a mistaken assumption about the probability of death among the chicks. These possibilities will be investigated presently.

The general statement of the theory is this: if there is a constant probability, p , in random subsamples of k , the relative frequency of 0, 1, 2, . . . occurrences is given by successive terms of the binomial

$$(q + p)^k,$$

where $q = 1 - p$ is the probability that the event will not occur; the expected frequencies in the classes of the sample are the products of these relative frequencies by n .

You may find it profitable to verify this theory by sampling a homogeneous population. Selecting a lot of beans of uniform size, shape and texture, color (or otherwise mark) some fraction of them, say one-third, then mix thoroughly. Draw subsamples of five, for example, by some random method of selecting one bean at a time, noting its color then replacing it among the others. Continue after re-mixing each time. The probability of getting 0, 1, 2, 3, 4, or 5 colored beans is taken from the successive terms of the expansion, $p = 1/3$, $q = 2/3$, $k = 5$:

$$\begin{aligned}
 \left(\frac{2}{3} + \frac{1}{3}\right)^5 &= \left(\frac{2}{3}\right)^5 + 5\left(\frac{2}{3}\right)^4\left(\frac{1}{3}\right) + 10\left(\frac{2}{3}\right)^3\left(\frac{1}{3}\right)^2 + \dots \\
 &= \frac{32}{243} + \frac{80}{243} + \frac{80}{243} + \frac{40}{243} + \frac{10}{243} + \frac{1}{243}
 \end{aligned}$$

If you carry on to 243 subsamples, $n = 243$, the expected frequencies will be the numerators of the six fractions.

Instead of beans, the random digits of table 1.2 may be used. The digits 3, 6, and 9 take the place of colored beans, while 1, 2, 4, 5, 7, and 8 stand for white: 0 is discarded when drawn. By some such convention, any desired probability may be specified.

The *binomial coefficients*, for the formation of which you will find several rules, are easily remembered by writing them down this way:

Size of Subsample k	Binomial Coefficients									
1										
2					1		2		1	
3				1		3		3		1
4			1		4		6		4	
5			1		5		10		10	
6			1		6		15		20	
7		1		7		21		35		21
8		1		8		28		56		28
etc.										

In this form, called *Pascal's triangle*, each coefficient is the sum of the two just above it to the right and left. The coefficients for $k = 4$ and $k = 5$ verify those used above for the chicks and the beans. You see, the distribution of the coefficients is always symmetrical. If $p = q = 0.5$ the binomial distribution is also symmetrical, the shape approximating more and more closely that of the normal distribution as k gets larger. If the probabilities are unequal, the binomial distribution is skewed.

In the binomial distribution, as in the normal, the mean and standard deviation are appropriate averages. Setting up the *expected numbers* in the usual manner for calculation (section 8.2):

Number of Living Chicks X	Expected Frequency f	fX	fX^2
0	6	0	0
1	24	24	24
2	36	72	144
3	24	72	216
4	6	24	96
Total	96	192	480

$$m = 192/96 = 2 \text{ chicks per lot}$$

$$\sigma = \sqrt{\frac{480 - (192)^2/96}{96}} = 1 \text{ chick per lot}$$

The symbols m and σ for mean and standard deviation are used to distinguish these population parameters from the sample statistics, $\bar{x}$ and s . The latter, calculated from the *observed* frequencies, are 2.2188 chicks per lot and 0.920 chick per lot: we shall have little use for these statistics. Another point you will notice is that for σ , division was by $n = 96$ instead

of by $n - 1 = 95$. Degrees of freedom are appropriate for computing sample statistics but not parameters.

One usually finds these averages expressed as proportions or as percentages rather than as actual numbers per lot. Thus, one reads that the proportion of living chicks expected is $m = 2/4 = 0.5$ with $\sigma = 0.25$; or that the percentage is 50% of living chicks with standard deviation 25%. This emancipates the averages from the subsample size, making those for lots of four directly comparable with similar ones for lots of any other size.

On account of mathematical relations connected with the binomial distribution, it is unnecessary to go through the calculations shown above. The parameters are got more easily from the formulas,

$$m = kp = (4)(0.5) = 2 \text{ chicks per lot}$$

$$\sigma = \sqrt{kpq} = \sqrt{(4)(0.5)(0.5)} = 1 \text{ chick per lot}$$

If the parameters are to be expressed as proportions or percentages, the formulas are

$$m = p = 0.5 = 50\%,$$

$$\sigma = \sqrt{pq/k} = \sqrt{(0.5)(0.5)/4} = 0.25 = 25\%$$

Thus, p is not only the probability of occurrence of an event but is also the population mean proportion or percentage of occurrences.

It is worth noticing that, if the subsample size were 16 chicks, the expected mean percentage would be the same, $p = 0.5 = 50\%$; whereas the expected standard deviation would be only,

$$\sigma = \sqrt{pq/k} = \sqrt{(0.5)(0.5)/16} = 0.125 = 12.5\%$$

The theoretical standard deviation, then, decreases as the subsample size increases. Really, σ is the standard error of the subsample expected mean, varying inversely as the square root of the subsample size—a familiar relation in measurement data.

In the whole binomially distributed sample the standard error of the mean is inversely proportional to the square root of the entire number of individuals. That is, if there are n subsamples of k individuals each,

$$\sigma_{\bar{x}} = \sqrt{\frac{pq}{nk}} = \frac{\sigma}{\sqrt{n}},$$

again a familiar relation. For the entire distribution of chicks, the standard error in percentage is

$$\sigma_{\bar{x}} = \frac{25}{\sqrt{96}} = 2.55\%$$

This question was proposed above: may we attribute to sampling variation the discrepancy between the assumed parameter, $m = 50\%$, and the corresponding sample mean, $\bar{x} = 2.2188/4 = 55.47\%$? Naturally,

if the sample is from a binomial distribution the test involves chi-square. Rather than use percentages, it is simpler to ask the question directly: is there a significant difference between the expected number of living chicks and the observed number? The former is 50% of the total number of chicks, 50% of 384 being 192 chicks. The latter is the value of $\sum fX$ in the sample,

$$\sum fX = (1)(0) + (20)(1) + (41)(2) + (25)(3) + (9)(4) = 213$$

In the usual manner,

$$\chi^2 = \frac{(213 - 192)^2}{192} + \frac{(171 - 192)^2}{192} = 4.59,$$

whose probability of occurrence is about 3.5%. The sample number of surviving chicks is, therefore, significantly greater than the proposed 50%.

It is common to find t used instead of χ^2 for the test. In the case under discussion, the tests are identical provided σ is used— s might lead to an erroneous conclusion. To illustrate, we found that the standard error of the mean was

$$\sigma_{\bar{x}} = 25/\sqrt{96} = 2.55\%$$

The difference between the sample mean and the parameter is

$$55.47 - 50 = 5.47\%$$

Then,

$$t = 5.47/2.55 = 2.15,$$

near its 3.5% level, $d.f. = \infty$, just as chi-square was near its 3.5% point. It is interesting to notice that, in these circumstances, $t = \chi$. That is the reason the t -test is valid for this kind of enumeration statistics.

EXAMPLE 16.1—In the manner of section 8.2 calculate $\bar{x}$ and s for the sample of surviving chicks.

EXAMPLE 16.2—For the proposed sampling of beans, $p = 1/3$, $k = 5$, show that $m = 1.667$ and $\sigma = 1.054$ colored beans per subsample, or that $m = 33.33\%$ with $\sigma = 21.08\%$.

EXAMPLE 16.3—It is convenient to bear in mind that the total number of individuals in a binomial sample is nk , and that the total number of occurrences of the event is $\sum fX$. Hence, $\bar{x} = 100(\sum fX)/(nk)$ percent.

EXAMPLE 16.4—If 243 subsamples of the beans ($k = 5$, $p = 1/3$) were drawn at random with $\bar{x} = 30.37\%$, show that $t^2 = \chi^2 = 4.80$. What conclusion about the sample would you reach?

EXAMPLE 16.5—You often see some such statement as this: "As the size of the sample is increased, the sample mean draws nearer and nearer to the population value." This would require chi-square to approach zero. Do you think it would?

As the sample increases in size, the binomial distribution approaches the normal, more rapidly for values of p near 0.5 than for those near 0 or 1. One consequence is that t can be used instead of table 1.1 for setting confi-

dence limits. If $n = 100$, for example, and the sample contains 40 individuals with the attribute being observed, then $p = 0.4$, $q = 0.6$, and

$$40 \pm t_{.05}\sqrt{npq} = 40 \pm 1.96\sqrt{(100)(0.4)(0.6)} = 40 \pm 9.6$$

specifies the 95% confidence interval, $30.4 - 49.6$. Compare this with the interval, $30 - 50$, given in the table.

As another example, suppose 100 individuals having a certain attribute are drawn in a sample of 1,000. Since in the corresponding part of table 1.1 ratios are being used, the interval is given by

$$0.1 \pm t_{.05}\sqrt{pq/n} = 0.1 \pm 1.96\sqrt{(0.1)(0.9)/1,000} = 0.1 \pm 0.0186$$

The interval, then, is from 0.0814 to 0.1186; that is, from 8.14% to 11.86%.

16.3—Comparison of a sample distribution with the binomial. In experimental data there is everpresent sampling variation to be considered. Under the assumption of a binomially distributed population as the source of the sample, we have already tested the significance of the deviation of the sample mean from m . The question now to be answered is whether the entire distribution of the sample departs from expected. The methods are identical with those described in chapter 9.

For illustration we use the data in table 16.1 (6). Since these are eight-pig litters, there are nine classes, having respectively 0, 1, . . . 8 males.

TABLE 16.1
EIGHT-PIG LITTERS HAVING VARIOUS NUMBERS OF MALES. EXPECTED NUMBERS, AND
TEST OF BINOMIAL DISTRIBUTION

Number Males in Litter X	Number of Litters f	Expected Number	Deviation, Number - Expected	(Deviation) ²	(Deviation) ² Expected
0	0	0.414	-0.414		
1	5	3.312	1.688		
2	9	11.594	-2.594		
3	22	23.188	-1.188		
4	25	28.984	-3.984		
5	26	23.188	2.812		
6	14	11.594	2.406		
7	4	3.312	0.688		
8	1	0.414	0.586		
Total	106	106.000	0.000		$\chi^2 = 1.948$

Let us first set up the hypothesis of equal probability of the sexes. Then $p = 0.5$, $q = 0.5$, $k = 8$, $n = 106$, and the expected frequencies are given by

$$\begin{aligned} &106 [(0.5)^8, 8(0.5)^7 (0.5), 28(0.5)^6 (0.5)^2, \text{etc.,}] \\ &= (106)(0.5)^8 [1, 8, 28, 56, 70, \text{etc.,}] \\ &= 0.414, 3.312, 11.594, \text{etc.} \end{aligned}$$

The deviations of these expected numbers from actual measure the departure of the sample distribution from the binomial being considered. Is the departure significant? The appropriate test is made by means of chi-square. But since each expected number should be at least as great as five, the first three and last three classes are pooled for the test. The resulting chi-square is moderate for $d.f. = 4$. Actually, there is somewhat less deviation from expected than would be found ordinarily in random sampling.

Why four degrees of freedom? The sum, 106 litters, calculated from the sample and used in the fitting, takes one degree. Effectively, there were only five classes used, leaving four degrees of freedom for the test.

From the results of the comparison, there is no evidence against the hypothesis of a uniform probability of 50% for maleness. This is, of course, not proof that the sex ratio is 1:1. It is merely that there is not a significant departure from that binomial distribution in which $p = q$.

EXAMPLE 16.6—Parkes (7) collected information about sex distribution from the National Duroc Jersey Pig Record. Among 402 eight-pig litters he recorded the numbers having 0, 1, 2, . . . 8 males as follows: 1, 8, 37, 81, 162, 77, 30, 5, 1. Compute $\chi^2 = 38.7$, $d.f. = 6$. Can you explain the significant deviation from the hypothetical distribution; that is, why the probability of maleness varies significantly among the litters?

EXAMPLE 16.7—Ignoring the nonbinomial distribution of the sample, test the significance of the difference between the expected number of males in the foregoing example, 1,608, and the number recorded in the Record, 1,581. Chi-square is 0.90. Does this throw any light on the sample distribution?

EXAMPLE 16.8—Test the significance of the deviation of the sample of chicks from the binomial with $p = 50\%$. Chi-square = 7.07, $d.f. = 4$. The mean, you may recall, differed significantly from 50%.

EXAMPLE 16.9—Test the significance of the deviation of the sample number of males (439) in table 16.1 from the number expected (424) under the 50% hypothesis. Chi-square = 1.06, $d.f. = 1$.

16.4—The test of homogeneity in a distribution of the binomial form. Often there is no theory upon which to base a hypothesis about the value of p . In that event the question raised and answered in the foregoing section has no meaning. Instead, we may wish to know whether the sample is *homogeneous*; that is, may it be a sample drawn from some binomial distribution, p unspecified? Since there is no theoretical probability, we simply take the sample mean percentage and set up a binomial distribution having that value of p . Naturally we get no test of the value of p , but we do learn if this probability may have been uniformly distributed over all the individuals.

The distribution of surviving chicks furnishes a good illustration. The evidence is strong that the hypothesis of 50% survival was not well founded (section 16.2, example 16.8). May the sample be a random drawing from

a binomial distribution with $p = 55.47\%$, the sample mean? The successive terms of the binomial,

$$96(0.4453 + 0.5547)^4$$

are the expected frequencies. These are easily calculated:

X	Powers of q	Powers of p	Binomial Coefficients	Relative Frequencies	Expected Frequencies
0	0.0393		1	0.0393	3.77
1	0.0883	0.5547	4	0.1959	18.81
2	0.1983	0.3077	6	0.3661	35.15
3	0.4453	0.1707	4	0.3040	29.18
4		0.0947	1	0.0947	9.09
Total				1.0000	96.00

Place $q = 0.4453$ in the last row but one, and in the calculating machine. Multiply repeatedly by q until q^* is reached, recording each power in the table. Repeat for p , entering the successive powers downward in the table beginning with the second row. In the fifth column each relative frequency is the product of the three (or two) numbers to its left. As an example, in the third row,

$$(0.1983)(0.3077)(6) = 0.3661$$

The relative frequencies must add to unity. Multiplication of each of them by $n = 96$ yields the binomially distributed frequencies for $p = 55.47\%$.

The chi-square test is shown in table 16.2. The degrees of freedom are now only two—not only does the sample total, $n = 96$, impose a limit on the expected frequencies, but also the sample mean has been made the parameter of the binomial. Nevertheless, chi-square exceeds its 50% point but little. There is no evidence against the constancy of

TABLE 16.2
TEST OF HOMOGENEITY OF THE CHICK SURVIVALS

X	f	Expected	Deviation	$\frac{(\text{Deviation})^2}{\text{Expected}}$
0	1	3.77		
1	20	18.81		
2	41	35.15		
3	25	29.18		
4	9	9.09		
	96	96.00	0.00	$\chi^2 = 1.68$

TABLE 16.3
NUMBER OF MALES IN LITTERS OF DUROC JERSEY PIGS

Five-pig Litters			Six-pig Litters		
Number Males in Litter	Number Litters	Expected Numbers	Number Males in Litter	Number Litters	Expected Numbers
0	2	4.08	0	3	3.62
1	20	19.46	1	16	21.37
2	41	37.07	2	53	52.55
3	35	35.31	3	78	69.08
4	14	16.84	4	53	50.98
5	4	3.21	5	18	20.11
			6	0	3.29
Total	116	$\chi^2 = 0.73$		221	$\chi^2 = 3.92$
	$p = 48.79\%$			$p = 49.62\%$	

the probability of survival. The population may be homogeneous, with a uniform probability of 55.47%.

After two samples have been found homogeneous, it is often desired to test the significance of the difference between their means. Assuming a uniform probability in each, do these probabilities differ significantly? In table 16.3 are the distributions of males in five- and six-pig litters. Each is homogeneous with respect to its own probability. Do these means differ significantly? The test is easily carried out in a 2×2 table like this:

Litters	Males	Females	Total
Five-pig	283	297	580
Six-pig	658	668	1,326
Total	941	965	1,906

$$\chi^2 = \frac{[(283)(668) - (297)(658)]^2(1906)}{(580)(1,326)(941)(965)} = 0.111$$

For $d.f. = 1$, this is a smaller value of chi-square than usual. The small difference between the two percentages,

$$49.62 - 48.79 = 0.83\%,$$

is not significant.

If a pooled value of p , $941/1,906 = 49.3704$, is used to compute the standard deviation, $\sqrt{pq(1/n_1 + 1/n_2)} = 0.0249$ this test may be carried through with t in the manner of the last section. That is, χ with one degree of freedom is the same as t with $d.f. = \infty$.

If there are three or more samples to be compared, the method of section 9.6 is probably the easiest.

What shall be done with samples that are not homogeneous? If there is no uniform probability of occurrence, and if nothing is known of the laws governing the probability, any conclusions are based on flimsy evidence. There is no assurance that, if the experiment were repeated, the same set of probabilities would be existent. If one doesn't know why the probabilities change or in what manner, he can set up no experimental controls. The first step must be toward improved knowledge of the techniques of selecting and handling the experimental material. Only after the sources of variation are discovered may valid comparisons be made.

Nevertheless, there are times when one wishes to make a test of significance of the difference between means despite heterogeneity. If means differ significantly even in the face of large variation, perhaps the difference represents some population fact. The method used to make the test must be sensitive to the variability present. Probably the best is to calculate the standard deviation of each set in the manner of section 8.2, then test significance as in section 8.8. The assumption of normal distribution there made is more flexible than the assumption of constant p .

You may find interest in contrasting the methods of treating the binomial distribution and the normal. Each has two parameters. A test of significance of sample deviation from theoretical distribution is available in each. But there is a single, basic normal distribution, the relative frequencies being set out in table 8.6, whereas there is a separate binomial for each p and k . It follows that tables of the binomial distributions would be clumsy—it is easier to compute each time from the parameters.

EXAMPLE 16.10—Calculate chi-square for each of the samples of table 16.3.

EXAMPLE 16.11—Decker and Andre (4) exposed adult chinch bugs at a temperature of -8.5°C . for 15 minutes, then counted the number dead in 100 subsamples of 10 bugs. The numbers of subsamples having 0, 1, 2, . . . 7 dead were 4, 21, 22, 28, 14, 8, 2, 1. No subsample had 8 or more dead. Is there a significant lack of homogeneity for probability of death? Chi-square = 2.89, df = 8.

EXAMPLE 16.12—The sample probability of maleness in example 16.6 is 49.160%. Is the sample homogeneous with respect to this percentage? Chi-square = 38.15, df = 5.

EXAMPLE 16.13—In example 16.9 the sample proportion of males in table 16.1 was found to be not significantly different from 0.5. Is the sample homogeneous with respect to the probability 0.5177? Chi-square = 0.73, df = 3.

EXAMPLE 16.14—Using the methods of chapter 8, test the significance of the difference between the mean percentages of males in the samples of table 16.1 and example 16.6. Since the latter departs significantly from binomial distribution, the use of chi-square in a 2×2 table is inappropriate. Answer, t = 0.35.

EXAMPLE 16.15—Test the significance of the difference between the percentages of males in the five- and six-pig litters of table 16.3, using t instead of χ^2 . Your result should be $t = \chi = \sqrt{0.111}$.

16.5—The Poisson distribution. Like the binomial, this is the distribution of a *discrete* or *discontinuous* variable arising from enumeration, usually only the integral values occurring. Also, there is the idea of the

uniform probability of the occurrence of some event, but this probability, ordinarily small, is somewhat ill defined. In fact, the size of the subsample, though supposed to be constant, is usually not known. It is large, however, and must be confined within reasonably narrow limits.

In contrast with both the binomial and the normal, the Poisson distribution is defined by a single parameter, the mean. The variance is equal to the mean.

The Poisson distribution can be shown to be a limiting form of the binomial in which p is very small (9). The distribution is valuable in its own right, however, quite independent of this relation to the binomial.

If m is the expected mean frequency of occurrence of some event in the population of subsamples, and if the frequency follows the Poisson distribution, then the event will occur 0, 1, 2, . . . times in subsamples with relative frequencies equal to the successive terms of this sequence,

$$\frac{1}{e^m}, \frac{m}{e^m}, \frac{m^2}{2e^m}, \frac{m^3}{(2)(3)e^m}, \frac{m^4}{(2)(3)(4)e^m}, \dots,$$

where e , the base of the natural logarithm, is an irrational number whose value is about 2.718, and whose logarithm (base 10) is 0.434295 very nearly. The expected frequencies are products of these relative frequencies by the number of subsamples, n .

The fitting of a Poisson distribution to a sample will be illustrated by the data on noxious weed seeds (5) in table 16.4. The subsamples are specified as weighing $\frac{1}{4}$ ounce. Naturally, the number of seeds in the subsample varies somewhat, but is large compared with the number of noxious weed seeds. The Poisson distribution, like the binomial, is built

TABLE 16.4
DISTRIBUTION OF 98 QUARTER-OUNCE SUBSAMPLES OF *Phleum Pratense* CLASSIFIED BY
NUMBER OF NOXIOUS WEED SEEDS. FITTED POISSON DISTRIBUTION.
TEST OF SIGNIFICANCE

Numbers of Noxious Weed Seeds	Number of Subsamples f	Expected Numbers	Deviations From Expected	(Deviations) ² Expected
0	3	4.78	-1.78	0.663
1	17	14.44	2.56	0.454
2	26	21.81	4.19	0.805
3	16	21.96	-5.96	1.618
4	18	16.58	1.42	0.122
5	9	10.02	-1.02	0.104
6	3	5.04		
7	5	2.18		
8	0	0.82		
9	1	0.28		
10		0.08		
More		0.01		
Total	98	98.00	0.00	$\chi^2 = 3.807$

on the assumption that the probability that any seed is that of a noxious weed is constant throughout the population. We shall first fit the theoretical distribution, then test the significance of the deviation of the sample.

The first step is the calculation of the sample mean in the manner of section 8.2:

$$\bar{x} = (SfX)/(Sf) = 296/98 = 3.02041 \text{ noxious weed seeds per subsample}$$

We then set up a Poisson distribution with this mean as its parameter, m . A distribution with some theoretical mean might be fitted, but that would be unusual. Ordinarily the sample mean is used. We are testing homogeneity only—is there a uniform probability of occurrence of weed seeds throughout the sample?

The remainder of the calculation below is by means of logarithms. The successive steps are rather evident. After the logarithms of the first two expected numbers are reached, those of the successive numbers are got by (i) adding the logarithm of m , and (ii) subtracting the logarithm of the successive integers beginning with 2. You can see this by writing the expected numbers in this way:

$$\frac{n}{e^m}, \frac{nm}{e^m}, \left(\frac{nm}{e^m}\right)\left(\frac{m}{2}\right), \left(\frac{nm^2}{2e^m}\right)\left(\frac{m}{3}\right), \dots$$

The calculations proceed thus:

Symbol	Logarithm	Expected Number
$n = 98$	1.99123	
e^m	$(3.02041)(0.434295) = 1.31175$	
n/e^m	0.67948	4.78
$m = 3.02041$	0.48007	
mn/e^m	1.15955	14.44
m	0.48007	
	1.63962	
2	0.30103	
$m^2n/2e^m$	1.33859	21.81
m	0.48007	
	1.81866	
3	0.47712	
$m^3n/(2)(3)e^m$	1.34154	21.96
m	0.48007	
	1.82161	
4	0.60206	
$m^4n/(2)(3)(4)e^m$	1.21955	16.58
m	0.48007	
	1.69962	
5	0.69897	
	1.00065	10.02

m	0.48007	
	<u>1.48072</u>	
6	0.77815	
	<u>0.70257</u>	5.04
m	0.48007	
	<u>1.18264</u>	
7	0.84510	
	<u>0.33754</u>	2.18
m	0.48007	
	<u>10.81761</u> - 10	
8	0.90309	
	<u>9.91452</u> - 10	0.82
m	0.48007	
	<u>10.39459</u> - 10	
9	0.95424	
	<u>9.44035</u> - 10	0.28
m	0.48007	
	<u>9.92042</u> - 10	
10	1.	
	<u>8.92042</u> - 10	0.08
Total		<u>97.99</u>

From the logarithm of 98 is subtracted the logarithm of e^m which is $m \log e = (3.02041)(0.434295)$. The difference is the logarithm of the quotient, the anti-logarithm being the expected number of subsamples containing no weed seeds. Next is added the logarithm of 3.02041 to get the logarithm of the next expected number, 14.44. Thereafter the logarithm of m is added each time while the logarithm of the next higher integer is subtracted. The process continues till the expected numbers become smaller than 5. We carried on to illustrate the process, but had to recombine for the test in table 16.4.

Chi-square is computed as usual. How many degrees of freedom are there? The sample mean has been set up as the parameter of the distribution, and the sample number of subsamples, 98, is used to calculate the expected numbers remaining after the fifth (or tenth) class. Hence two degrees of freedom must be deducted from the effective number of classes, leaving $df. = 5$ for testing chi-square. Clearly, the sample does not depart significantly from the Poisson distribution. The 98 bags of seeds from which the subsamples were taken may constitute a population with Poisson distribution. The probability that any randomly selected seed be that of a noxious weed may be uniform throughout.

The variance in the theoretical distribution is the same as the mean, 3.0204. Hence $s = \sqrt{3.0204} = 1.738$ seeds per subsample. Obviously, the standard deviation has different relations to the Poisson distribution than to the normal. For example, it cannot be said that 68.27% of the

individuals in the unsymmetrical Poisson distribution lie within the interval, $m \pm \sigma$. However, if the entire sample of 98 quarter ounces be considered as a unit with 296 weed seeds, it may be thought of as a single sample drawn from a population whose mean is estimated as 296 and whose variance is therefore 296. Poisson populations with such large means are distributed almost normally. Hence, for practical purposes this large sample may be treated as though drawn from a normal population whose mean is estimated as $\bar{x} = 296$ with standard error, $s_{\bar{x}} = \sqrt{296} = 17.2$ weed seeds. The numbers of weed seeds in two such samples may then be compared, the difference being tested for significance in the manner of the next section.

EXAMPLE 16.16—Leggatt counted various other weed seeds in his 98 subsamples. Numbers of two genera follow:

Genus	Number of Seeds in Subsample								Chi-square	P
	0	1	2	3	4	5	6	7		
Potentilla	37	32	16	9	2	0	1	1	2.931	24%
Carex	36	36	23	2	1				4.476	11%

EXAMPLE 16.17—The chinch bug data of example 16.11 are well adapted to binomial fitting. It may be interesting to apply the Poisson distribution. Chi-square = 4.634, *d.f.* = 5.

EXAMPLE 16.18—If you have the book available, look at table 51 in Pearson's (8) collection and compare the distribution for $m = 3.0$ with that of table 16.4. Make a rough check on your work in example 16.17 by use of the same table.

EXAMPLE 16.19—If m is large, the Poisson distribution is approximately normal. To verify this, work out the distribution for $m = 15$, or take it from Pearson's table. Group the expected frequencies for $n = 100$ as follows:

Class mark	3	6	9	12	15	18	21	24	27
Frequency	1	1	10	24	30	22	9	2	1

Using the method of section 8.6, show that $g_1 = 0.094 \pm 0.24$ and $g_2 = 0.46 \pm 0.48$.

EXAMPLE 16.20—The sample of table 16.4 may be looked upon as a single subsample of 24.5 ounces drawn from a Poisson population with $m = 296$ noxious weed seeds. The variance of this mean is 296. Comparing this with the foregoing example, you see why large means from Poisson populations may be considered as normally distributed.

16.6—Comparison of means in samples from Poisson distributions. A moment's thought will make it clear that means from Poisson distributions cannot be compared like those from the binomial. The records give the number of times the event occurs but not the times it fails. For small samples a chi-square method of testing was explained in section 9.12. But since large sample means drawn from a Poisson population are distributed

almost normally, it turns out that the appropriate method of testing differences is that of section 8.8. An example will recall the method.

"Student" (9) gave two distributions of yeast cells counted in the 400 squares of a haemocytometer, table 16.5. The values of chi-square show

TABLE 16.5
TWO SAMPLES OF YEAST CELLS COUNTED IN 400 SQUARES OF A HAEMACYTOMETER
Fitted Poisson Distributions

Sample	Number Squares Having Specified Number of Cells										Chi-square	
	0	1	2	3	4	5	6	7	8	9	Mean	square
1	Observed	103	143	98	42	8	4	2			529	3.177
	Expected	106	141	93	41	14	5					
2	Observed	75	103	121	54	30	13	2	1	0	720	7.073
	Expected	66	119	107	64	29	15					

that they may well be samples from Poisson distributions. The difference between the means is $720 - 529 = 191$ yeast cells per sample. To test significance, we have the two variances, the same as the means, 720 and 529. From these, the standard error of the mean difference is

$$\sqrt{720 + 529} = 35.3$$

Hence, $t = 191/35.3 = 5.41$, strong evidence that the two samples are not drawn from the same population.

EXAMPLE 16.21—Calculate the expected numbers in table 16.5, together with the values of chi-square.

EXAMPLE 16.22—"Student" recorded one sample of yeast cells with this distribution:

Number of yeast cells	0	1	2	3	4	5
Number of squares	213	128	37	18	3	1

Compute $\chi^2 = 9.683$, $d.f. = 2$. This departs significantly from Poisson distribution.

EXAMPLE 16.23—Using the methods of chapter 8, test the significance of the difference between the means of the two samples in table 16.5. You would be assuming that mean differences are normally distributed, but you would avoid the formula, $\bar{x} = s^2$. Although this formula is justified in the samples of table 16.5, it would not be appropriate if the samples departed significantly from Poisson distribution. Ans., $t = 5.30$.

EXAMPLE 16.24—In table 16.5, carry through the test of significance by comparing the mean numbers of yeast cells *per square*, 1.3225 and 1.800. You should reach the same value of t .

16.7—Transformations of data for test of significance. In a binomial distribution the variance, kpq or $kp(1 - p)$, is clearly related to the mean, kp ; while in the Poisson the two are equal. In measurement data, also, it

is by no means unusual to come across groups in which similar relations appear, or in which the mean and standard deviation are proportional. If tests of significance by use of t or F are to be applied, it may be desirable to transform such variates into new sets having the mean and variance independent (section 11.17). To meet the situations which ordinarily occur suitable transformations have been devised (1) (2).

Consider first an investigation in which the occurrence of some event is being counted, as for example, the number of wireworms found in an experimental plot or under a square foot of sod. If the counts run 50 or more per plot, they may be analyzed usually without change. Examine the ranges in the several treatments, however, to detect any tendency for the variance to be proportional to the mean. Counts of less than 50 per plot are likely to be noticeably distributed in the Poisson fashion. The appropriate transformation is the square root of the count. Make a new table in which the square root of the number of wireworms is entered instead of the actual count, then proceed as usual.

In experiments where 10 or less are counted in the plots, the square root tends to over-correct, but $\sqrt{X + 0.5}$ may be used, X being the observed number. This method was followed in testing the data of table 16.6 (3).

TABLE 16.6

NUMBER OF WIREWORMS IN THE PLOTS OF A LATIN SQUARE, THE TREATMENTS BEING A CHECK (O) AND FOUR DIFFERENT SOIL FUMIGATIONS APPLIED IN THE PREVIOUS YEAR

Rows	Columns				
	1	2	3	4	5
Treatment Wireworms, X	P	O	N	K	M
$\sqrt{X + 0.5}$	3	2	5	1	4
	1.87	1.58	2.34	1.22	2.12
Treatment Wireworms, X	M	K	O	N	P
$\sqrt{X + 0.5}$	6	0	6	4	4
	2.55	0.71	2.55	2.12	2.12
Treatment Wireworms, X	O	M	K	P	N
$\sqrt{X + 0.5}$	4	9	1	6	5
	2.12	3.08	1.22	2.55	2.34
Treatment Wireworms, X	N	P	M	O	K
$\sqrt{X + 0.5}$	17	8	8	9	0
	4.18	2.92	2.92	3.08	0.71
Treatment Wireworms, X	K	N	P	M	O
$\sqrt{X + 0.5}$	4	4	2	4	8
	2.12	2.12	1.58	2.12	2.92
Treatments Means of square roots	O	N	K	P	M
	2.45	2.62	1.20	2.21	2.56

Analysis of Variance

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Rows	4	2.4812	
Columns	4	0.9026	
Treatments	4	6.8747	1.719
Error	12	4.4551	0.371

$$F = 1.719/0.371 = 4.63^*$$

It is clear that K was the treatment responsible for significance. In applying the methods of chapter 15, use the sums and sums of squares of the square roots, not of the original numbers. But in presenting the results, the means should be changed back to numbers of wireworms. Thus, the appropriate mean number for O is

$$(2.45)^2 - 0.5 = 5.50 \text{ wireworms per plot}$$

Do not attempt to apply any single standard error to the resulting mean numbers because each has its own. Carry out tests of significance with the transformed means.

To acquire judgment about transformations, analyze the original data and compare the two results.

The second transformation to be presented is one appropriate to the kind of data associated with the binomial—so many occurrences among a definite total counted, or percentage of occurrences. For illustration we use the data in table 16.7, courteously provided by Dr. J. C. Gaines of the Texas Agricultural Experiment Station. The counts of punctured squares were first converted to percentages. Since the numbers of punctured squares were all less than 100, a transformation is advisable. The one to be used converts percentages into angles,

$$\text{angle} = \arcsin \sqrt{\text{percentage}}$$

For convenience we give this transformation in table 16.8. As an example of its operation, let us get out the angle for 17.4%. Look for 17 in the left hand column and for 4 in the top row, reading the angle 24.65° . Again, take the percentage 0.56. In row 0.5 and column 6 read 4.29° .

If most or all of your percentages are above 50%, you may prefer to work with the other of the two conjoint events (*free* instead of *infested*, for example).

The analysis of variance of the angles shows that the treatments differed in their control of the boll weevil. However, treatments A , B , and C (different concentrations of calcium arsenate) had about equal effects, while D (calcium arsenate, Paris green and sulphur) was almost as ineffec-

tive as no treatment, *E*. The average percentages of squares punctured are 14.0, 12.3, 15.4, 24.7, and 30.6, easily read from table 16.8 corresponding to the mean angles shown in table 16.7.

It should be observed that conclusions would have been the same had the original data been used without transformation.

The third (and last) case to be discussed is that in which the group means and standard deviations tend to be proportional—the coefficients of variation are much the same. The proper transformation is to logarithms whether the data are counts, percentages or measurements. The method is illustrated by the catch of plankton nets at the left side of

TABLE 16.7

PERCENTAGE OF SQUARES PUNCTURED BY BOLL WEAVILS ON UNTREATED COTTON PLANTS (*E*) AND ON PLANTS DUSTED WITH FOUR MATERIALS. LATIN SQUARE LAYOUT. TWO HUNDRED SQUARES PER PLOT EXAMINED. ALSO, CORRESPONDING ANGLES IN DEGREES

Rows	Columns				
	1	2	3	4	5
Treatment 1	<i>C</i>	<i>E</i>	<i>A</i>	<i>D</i>	<i>B</i>
Percentage	27	42	18	34	17
Angle	31.3	40.4	25.1	35.7	24.4
Treatment 2	<i>A</i>	<i>D</i>	<i>E</i>	<i>B</i>	<i>C</i>
Percentage	18	27	42	14	12
Angle	25.1	31.3	40.4	22.0	20.3
Treatment 3	<i>D</i>	<i>B</i>	<i>C</i>	<i>E</i>	<i>A</i>
Percentage	23	14	17	25	14
Angle	28.7	22.0	24.4	30.0	22.0
Treatment 4	<i>E</i>	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>
Percentage	24	10	8	12	26
Angle	29.3	18.4	16.4	20.3	30.7
Treatment 5	<i>B</i>	<i>C</i>	<i>D</i>	<i>A</i>	<i>E</i>
Percentage	9	11	15	11	22
Angle	17.5	19.4	22.8	19.4	28.0
Treatment Mean angle	<i>A</i> 22.0	<i>B</i> 20.5	<i>C</i> 23.1	<i>D</i> 29.8	<i>E</i> 33.6

Analysis of Variance of Angles

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Rows	4	311.3	77.82
Columns	4	6.0	1.50
Treatments	4	637.5	159.38**
Error	12	108.8	9.07

TABLE 16.8

ANGLES CORRESPONDING TO PERCENTAGES, $\text{ANGLE} = \text{ARC SIN} \sqrt{\text{PERCENTAGE}}$,
AS GIVEN BY C. I. BLISS (2)

%	0	1	2	3	4	5	6	7	8	9
0.0	0	0.57	0.81	0.99	1.15	1.28	1.40	1.52	1.62	1.72
0.1	1.81	1.90	1.99	2.07	2.14	2.22	2.29	2.36	2.43	2.50
0.2	2.56	2.63	2.69	2.75	2.81	2.87	2.92	2.98	3.03	3.09
0.3	3.14	3.19	3.24	3.29	3.34	3.39	3.44	3.49	3.53	3.58
0.4	3.63	3.67	3.72	3.76	3.80	3.85	3.89	3.93	3.97	4.01
0.5	4.05	4.09	4.13	4.17	4.21	4.25	4.29	4.33	4.37	4.40
0.6	4.44	4.48	4.52	4.55	4.59	4.62	4.66	4.69	4.73	4.76
0.7	4.80	4.83	4.87	4.90	4.93	4.97	5.00	5.03	5.07	5.10
0.8	5.13	5.16	5.20	5.23	5.26	5.29	5.32	5.35	5.38	5.41
0.9	5.44	5.47	5.50	5.53	5.56	5.59	5.62	5.65	5.68	5.71
1	5.74	6.02	6.29	6.55	6.80	7.04	7.27	7.49	7.71	7.92
2	8.13	8.33	8.53	8.72	8.91	9.10	9.28	9.46	9.63	9.81
3	9.98	10.14	10.31	10.47	10.63	10.78	10.94	11.09	11.24	11.39
4	11.54	11.68	11.83	11.97	12.11	12.25	12.39	12.52	12.66	12.79
5	12.92	13.05	13.18	13.31	13.44	13.56	13.69	13.81	13.94	14.06
6	14.18	14.30	14.42	14.54	14.65	14.77	14.89	15.00	15.12	15.23
7	15.34	15.45	15.56	15.68	15.79	15.89	16.00	16.11	16.22	16.32
8	16.43	16.54	16.64	16.74	16.85	16.95	17.05	17.16	17.26	17.36
9	17.46	17.56	17.66	17.76	17.85	17.95	18.05	18.15	18.24	18.34
10	18.44	18.53	18.63	18.72	18.81	18.91	19.00	19.09	19.19	19.28
11	19.37	19.46	19.55	19.64	19.73	19.82	19.91	20.00	20.09	20.18
12	20.27	20.36	20.44	20.53	20.62	20.70	20.79	20.88	20.96	21.05
13	21.13	21.22	21.30	21.39	21.47	21.56	21.64	21.72	21.81	21.89
14	21.97	22.06	22.14	22.22	22.30	22.38	22.46	22.55	22.63	22.71
15	22.79	22.87	22.95	23.03	23.11	23.19	23.26	23.34	23.42	23.50
16	23.58	23.66	23.73	23.81	23.89	23.97	24.04	24.12	24.20	24.27
17	24.35	24.43	24.50	24.58	24.65	24.73	24.80	24.88	24.95	25.03
18	25.10	25.18	25.25	25.33	25.40	25.48	25.55	25.62	25.70	25.77
19	25.84	25.92	25.99	26.06	26.13	26.21	26.28	26.35	26.42	26.49
20	26.56	26.64	26.71	26.78	26.85	26.92	26.99	27.06	27.13	27.20
21	27.28	27.35	27.42	27.49	27.56	27.63	27.69	27.76	27.83	27.90
22	27.97	28.04	28.11	28.18	28.25	28.32	28.38	28.45	28.52	28.59
23	28.66	28.73	28.79	28.86	28.93	29.00	29.06	29.13	29.20	29.27
24	29.33	29.40	29.47	29.53	29.60	29.67	29.73	29.80	29.87	29.93
25	30.00	30.07	30.13	30.20	30.26	30.33	30.40	30.46	30.53	30.59
26	30.66	30.72	30.79	30.85	30.92	30.98	31.05	31.11	31.18	31.24
27	31.31	31.37	31.44	31.50	31.56	31.63	31.69	31.76	31.82	31.88
28	31.95	32.01	32.08	32.14	32.20	32.27	32.33	32.39	32.46	32.52
29	32.58	32.65	32.71	32.77	32.83	32.90	32.96	33.02	33.09	33.15
30	33.21	33.27	33.34	33.40	33.46	33.52	33.58	33.65	33.71	33.77
31	33.83	33.89	33.96	34.02	34.08	34.14	34.20	34.27	34.33	34.39
32	34.45	34.51	34.57	34.63	34.70	34.76	34.82	34.88	34.94	35.00
33	35.06	35.12	35.18	35.24	35.30	35.37	35.43	35.49	35.55	35.61
34	35.67	35.73	35.79	35.85	35.91	35.97	36.03	36.09	36.15	36.21
35	36.27	36.33	36.39	36.45	36.51	36.57	36.63	36.69	36.75	36.81
36	36.87	36.93	36.99	37.05	37.11	37.17	37.23	37.29	37.35	37.41
37	37.47	37.52	37.58	37.64	37.70	37.76	37.82	37.88	37.94	38.00
38	38.06	38.12	38.17	38.23	38.29	38.35	38.41	38.47	38.53	38.59
39	38.65	38.70	38.76	38.82	38.88	38.94	39.00	39.06	39.11	39.17
40	39.23	39.29	39.35	39.41	39.47	39.52	39.58	39.64	39.70	39.76
41	39.82	39.87	39.93	39.99	40.05	40.11	40.16	40.22	40.28	40.34
42	40.40	40.46	40.51	40.57	40.63	40.69	40.74	40.80	40.86	40.92
43	40.98	41.03	41.09	41.15	41.21	41.27	41.32	41.38	41.44	41.50
44	41.55	41.61	41.67	41.73	41.78	41.84	41.90	41.96	42.02	42.07
45	42.13	42.19	42.25	42.30	42.36	42.42	42.48	42.53	42.59	42.65
46	42.71	42.76	42.82	42.88	42.94	42.99	43.05	43.11	43.17	43.22
47	43.28	43.34	43.39	43.45	43.51	43.57	43.62	43.68	43.74	43.80
48	43.85	43.91	43.97	44.03	44.08	44.14	44.20	44.25	44.31	44.37
49	44.43	44.48	44.54	44.60	44.66	44.71	44.77	44.83	44.89	44.94

(Continued next page)

TABLE 16.9
ESTIMATED NUMBERS OF FOUR KINDS OF PLANKTON CAUGHT IN SIX HAULS
WITH EACH OF TWO NETS

Count	Estimated Numbers				Logarithms			
	I	II	III	IV	I	I	III	IV
1	895	1,520	43,300	11,000	2.95	3.18	4.64	4.04
2	540	1,610	32,800	8,600	2.73	3.21	4.52	3.93
3	1,020	1,900	28,800	8,260	3.01	3.28	4.46	3.92
4	470	1,350	34,600	9,830	2.67	3.13	4.54	3.99
5	428	980	27,800	7,600	2.63	2.99	4.44	3.88
6	620	1,710	32,800	9,650	2.79	3.23	4.52	3.98
7	760	1,930	28,100	8,900	2.88	3.29	4.45	3.95
8	537	1,960	18,900	6,060	2.73	3.29	4.28	3.78
9	845	1,840	31,400	10,200	2.93	3.26	4.50	4.01
10	1,050	2,410	39,500	15,500	3.02	3.38	4.60	4.19
11	387	1,520	29,000	9,250	2.59	3.18	4.46	3.97
12	497	1,685	22,300	7,900	2.70	3.23	4.35	3.90
Mean	671	1,701	30,775	9,396	2.802	3.221	4.480	3.962
Range	663	1,480	24,400	9,440	0.43	0.39	0.36	0.41

Analysis of Variance of Logarithms

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Kind of plankton	3	20.2070	6.7357
Haul	11	0.3387	0.0308
Discrepance	33	0.2300	0.0070

table 16.9 (10). You will observe the pronounced regression of mean on range. The corresponding logarithms have approximately equal ranges, clearly independent of the means.

The appropriate means for the four kinds of plankton (i.e., the *geometric means*) are the antilogarithms of 2.802, etc.; that is, 634, 1,663, 30,200, and 9,162.

The standard deviation of the logarithms is $\sqrt{0.0070} = 0.0837$, that of the numbers being the antilogarithm, 1.21. Quoting Winsor and Clarke (page 5), "Now a deviation of 0.0837 in the logarithm of the catch means that the catch has been multiplied (or divided) by 1.21. Hence we may say that one standard deviation in the logarithm corresponds to a percentage standard deviation, or coefficient of variation, of 21% in the catch."

REFERENCES

1. M. S. BARTLETT. *Journal of the Royal Statistical Society Supplement*, 3:68 (1936).
2. C. I. BLISS. *Plant Protection*, No. 12, Leningrad (1937).
3. W. G. COCHRAN. *The Empire Journal of Experimental Agriculture*, 6:157 (1938).

4. GEORGE C. DECKER and FLOYD ANDRE. Iowa State College Journal of Science, 10:403 (1936).
5. C. W. LEGGATT. Comptes rendus de l'Association Internationale d'Essais de Semences, 5:27 (1935).
6. HUGH C. MCPHEE. Journal of Agricultural Research, 34:715 (1927).
7. A. S. PARKES. Biometrika, 15:373 (1923).
8. KARL PEARSON. Tables for Statisticians and Biometricians, University Press, Cambridge, England (1924-1931).
9. "Student." Biometrika, 5:351 (1907).
10. C. P. WINSOR and G. L. CLARKE. Sears Foundation: Journal of Marine Research, 3:1 (1940).

Chapter 17

DESIGN AND ANALYSIS OF SAMPLINGS

17.1—Introduction. Having read the preceding chapters, you will readily agree that sampling is the central problem in statistics. From some population a sample of individuals is drawn. This sample is examined in detail, such statistics as mean, standard deviation, and regression coefficients being calculated. From these sample facts inductions about the population parameters (tests of significance and fiducial limits) are made. That is the routine procedure which you have been following throughout this book.

In chapters 1 and 3, having actually drawn a large number of samples you verified some of the better known mathematical theorems about them—the standard error of the mean, the distributions of χ^2 and t , etc. This led you to place confidence in other distribution theory presented in later chapters.

Finally, you have been provided with a number of the more common experimental designs, incorporating accepted principles of sampling. Well conducted experiments with standard designs may be expected to yield reliable information about the sampled populations.

We now come to a discussion of some rather general problems in sampling, with particular applications to samplings made by interview, by questionnaire, or by examination of limited portions of extensive areas. It will be clear, however, that the principles involved are in no wise different from those which have governed the methods already presented.

17.2—Populations. Specification of a population is not always easy. Mathematically, it is only necessary to fix certain parameters. As an example, its mean and standard deviation completely determine a normal population. But what population are you sampling when you do a randomized block experiment on corn yield? We think of this population as distributed in normal or near-normal fashion, and consisting of all yields of the varieties being tested wherever they might be planted on similar soil and with similar environmental conditions. It may even be that a population of this kind is wholly imaginary, as when we are testing some newly developed hybrid of which there is no surplus of seed, so that only the sample exists. Populations like these, indefinite in extent and location, need careful description. The investigator specifies them by giving details of his experimental conditions—soil type, fertilization, previous manage-

ment, and seasonal characteristics such as humidity, temperature, etc. An example of a somewhat more definite population is that of four-year-old Guernsey milk-cows in Waukesha County, Wisconsin. The individuals of this kind of population, finite in number and confined to a limited area, may conceivably be enumerated in census fashion.

These instances illustrate the variety of concepts that may be associated with the word, population. A more extensive discussion may be found in (11). Before undertaking a sampling, one must clearly delimit the population about which he expects to seek information. After the sample is summarized, he must be careful to localize his conclusions to the population actually sampled.

A population is said to be homogeneous if it has a single set of parameters. For example, the death rate of 30-year-old native born white males in Iowa would constitute a homogeneous population if every member of it were subject to the same probability of death, p . Again, farm income from a given size and type of owner farm in the Cash Grain District would be a homogeneous population if its distribution could be described by some such set of parameters as mean, standard deviation, and skewness. Clearly, the exact description of homogeneous populations is conceptual. These descriptions form useful models, however, to which actual populations often approximate.

The experimenter usually sets up controlled conditions, or describes actual conditions, thus limiting his populations to approximate homogeneity. But the populations with which the economist, the sociologist, and the crop estimator deal are usually heterogeneous. No one set of parameters describes any extensive group of individuals. Not only do the populations change with position and with time, but many populations may occupy the same area at the same time. These investigators attempt to segregate their various populations into subdivisions or *strata*, each stratum approximating homogeneity. This device of stratification has broad implications, some of which will be discussed below.

17.3—Sampling from a homogeneous population. The objective of sampling consists in the group of inferences which can be made about the population sampled. Obviously such inferences will be valuable only insofar as the sample is *representative* of the population. In this book a deal of attention has been given to the manner in which the sample is representative—the estimates it furnishes of population parameters and the probability statements that may be made about significance and fiducial limits. Now we wish to inquire the circumstances which insure representativeness.

People have different ideas of the restrictions imposed by the term, *representative*, because they approach sampling with different objects in view. Let us distinguish three objectives. The first is to get a sample

which will yield an unbiased estimate of the population mean. This end is attained if all members of the population have the same chance of being drawn into the sample. As an example we may suppose that there is a list of all the inhabitants and that we wish a sample of five per cent of them. Starting with any item from the first to the twentieth in the list, this and each twentieth item thereafter may be taken to make up the sample. To guarantee equal opportunity to all, it is only necessary to leave the choice of the initial item to some objective device such as drawing a numbered bean from a bag. If the bean marked 13 is drawn, then the sample is to consist of items numbered 13, 33, 53, etc. A similar plan is the grid often laid down on a city map, every tenth dwelling or family along each fifth street, say, being visited. Care must be taken, of course, to avoid multiples of the average number of dwellings on a block. With obvious minor precautions, samples drawn in such manner furnish the estimate desired. Clearly a sample so planned has the virtue of being spread uniformly over the list or area.

Another objective in sampling is that the estimates shall be as accurate as possible for the time and money spent. It is possible that a nonrandom sampling may accomplish this while a random sample would be less accurate. In the hands of experienced investigators with intimate knowledge of the population a selected sample may give good approximations to the population parameters. But the hazards in such sampling are far greater than those where chance has free play. One reason is that human choice is too often biased, more particularly in respect of variation (10) (2). Another is that all probability theory rests on the laws of chance so that in default of random selection one cannot safeguard his findings by knowledge of significance and fiducial limits. The disinterested scientist, wishing to make his samplings as nearly objective as possible, usually prefers to avoid all possibility of influencing his findings by selection of the individuals upon which they are based, and at the same time to take advantage of the internal evidence about variation which is inherent in random samples.

Still a third objective in sampling is one which has been rather persistently presented in this book, that the sample shall furnish an unbiased estimate of variance as well as of the mean, thus making possible tests of significance and fiducial estimates. To reach this objective we may specify that, in case samples of size n have been decided upon, every possible sample of n individuals shall have the same chance of being drawn. This condition is clearly not met in taking every twentieth item from a long list because there are only 20 such samples available out of the multitude possible. The condition is met, however, by drawing at random n individuals from the population. The identity of the two concepts is shown as follows: imagine drawing samples of three from a population with six beans, $A, B, \dots F$. The theory of probability as developed in

any text on algebra leads to the conclusion that there are 20 different samples which are possible,

<i>ABC</i>	<i>BCD</i>	<i>CDE</i>	<i>DEF</i>
<i>ABD</i>	<i>BCE</i>	<i>CDF</i>	
<i>ABE</i>	<i>BCF</i>	<i>CEF</i>	
<i>ABF</i>	<i>BDE</i>		
<i>ACD</i>	<i>BDF</i>		
<i>ACE</i>	<i>BEF</i>		
<i>ACF</i>			
<i>ADE</i>			
<i>ADF</i>			
<i>AEF</i>			

The probability of drawing any one of these samples is 0.05. Having drawn at random three beans without replacement, you find yourself in possession of some one of these triplets, thus satisfying the requirement of equal likelihood for all samples. Replacement is not necessary to insure equal chance to every sample. If it were, one would not be able to divide a group of pigs at random into lots of n .

Care is necessary in describing the population being sampled, and in affording equal chance of selection to every individual or to every sample. A sample drawn in random manner from a list of telephone subscribers is representative of the population of subscribers whose names are included in the list, but is unlikely to be representative of voters, of taxpayers, or of all inhabitants. People responding to a mailed questionnaire usually constitute a selective group among those receiving it: those who from habit or from oversight do not respond cannot be said to have any chance at all of appearing in the sample despite the fact that the questionnaire was sent to them. For the same reason, housewives who are not found at home by an interviewer must be called upon till found if a representative sample is to be got. These examples illustrate the pitfalls that must be avoided in sampling.

EXAMPLE 17.1—If a sample of 10 is drawn without replacement from a limited population of 100 individuals, is the probability of being drawn the same for every individual?

EXAMPLE 17.2—How many different samples of four individuals can be drawn from a population of 10? What is the probability of drawing one of them?

17.4—Size of sample. The designer of a sampling investigation is always confronted with a decision about the size of the sample. While resources of money and personnel often set limits which must not be passed, there is usually opportunity to take advantage of any basis of judgment that may be available.

Investigators are often content with samples less than 0.1% of the population, while a sample of 10% is usually considered large. Such loose statements can be replaced by more definite ones only if there is available

some knowledge of the mean and standard deviation of the population to be sampled. Although this knowledge may not be precise it enables us to make fair approximations to sample size.

We sometimes start with information about the standard deviation of a previous sample of n . If the sample was large, there may be little lack of precision in assuming that another sample will produce the same s . Take $\bar{x}$ as the estimate of m in the proposed sample, then the fiducial limits are specified by

$$\frac{\bar{x} - m}{s/\sqrt{n}} = t$$

Solving for n ,

$$n = t^2 s^2 / (\bar{x} - m)^2$$

We may now ask this question: how large a sample will be required to fix the fiducial limits at l units from the sample mean, where $l = \bar{x} - m$?

An example is found in an experimental sample to estimate the yield per acre of wheat in certain districts of North Dakota in 1938 (4). The variance of 222 fields was 90.3. For so large a sample, $t = 2.6$ may be taken as a safe approximation to the 1% level. Suppose we wish to say, after the sample is taken, that the probability is 0.99 that the true mean is within $l = 2$ bushels per acre of the sample mean. Then

$$n = (2.6)^2(90.3)/2^2 = 153 \text{ fields}$$

is a guide to the required sample size.

In the absence of knowledge about the variance of an earlier sample, estimates may be made in the manner of sections 5.7 and 5.8. One may know, for example, that the percentage of farm land in crops in a certain region varies from 30 to 80. From this it may be inferred that

$$s = (80 - 30)/6 = 8\%, \text{ and hence, } s^2 = 64$$

How many farms in this region should be included in a random sample to get an estimate of percentage farm land in crops within 1 point of the population mean?

$$n = (2.6)^2(64)/1^2 = 433 \text{ farms}$$

If it is the coefficient of variation about which you have knowledge, the fiducial interval should be expressed as a percentage of the mean. The question may be asked: what sample size may be expected to indicate m not further away from $\bar{x}$ than $p\%$ of $\bar{x}$? That is

$$\bar{x} - m = p\bar{x}/100$$

Substituting in the formula for n ,

$$n = (100)^2 t^2 s^2 / p^2 \bar{x}^2$$

But the coefficient of variation is given by

$$C^2 = (100)^2 s^2 / \bar{x}^2$$

whence,

$$n = t^2 C^2 / p^2$$

As an example, if it is known that the coefficient of variation for carcass weight of swine in a certain class is $C = 5\%$, we may inquire the number of randomly selected carcasses necessary to approximate the carcass mean within $p = 3\%$:

$$n = (2.6)^2 5^2 / 3^2 = 19 \text{ carcasses}$$

In such estimates, 2.6 is suggested for t as a safeguard against eventual sampling variation. The actual sample may turn up with unusual mean and standard deviation. To be on the safe side, somewhat larger samples than those calculated above should be drawn, especially if the indicated size is small. However, if you can risk a little less accuracy than you hope for, you may use the estimates of n with considerable confidence.

17.5—Sampling from small populations. There are instances in which the sample comprises 5% or more of some limited population, as when percentage of farm land in corn is sampled from a county with a section of land as the sampling unit. For example, suppose a sample of $n = 64$ section percentages has been drawn at random from a county whose population of such units is $N = 576$. The standard deviation of the sample calculated in the usual way is, say, $s = 10\%$. It seems reasonable that the possibilities of variation in sample means are more and more restricted as the sample size approaches that of the population. The formula for standard error appropriate for means drawn from a finite population is modified to (3)

$$s_s = \frac{s}{\sqrt{n}} \sqrt{1 - \frac{n}{N}}$$

which reduces to the usual form as N increases toward infinity. Substituting $s = 10$, $n = 64$, $N = 576$,

$$s_s = \frac{10}{\sqrt{64}} \sqrt{1 - \frac{64}{576}} = (1.25)(0.943) = 1.18\%,$$

instead of the value, 1.25%, which would have been calculated for the same sample from an unlimited population.

Adjustment should also be made in the binomial formula of section 16.2,

$$s = \sqrt{pq/n},$$

if objects are drawn without replacement from a limited group. Suppose 50 families in a community of 500 are asked by telephone if they are listening to a certain radio program, 10 replying "yes." Then, $p = 0.2$,

$q = 0.8$, and the best estimate of standard error of the mean proportion $p = 0.2$, is

$$s = \sqrt{\frac{pq}{n}} \sqrt{1 - \frac{n}{N}} = \sqrt{(0.0032)} \sqrt{0.9} = 0.053,$$

instead of $\sqrt{0.0032} = 0.057$.

EXAMPLE 17.3—If a sample of four from the 16 townships of a county has the standard deviation 45, show that the standard deviation of the mean is 19.5.

EXAMPLE 17.4—Following is the analysis of variance for an assumed sample of Iowa corn condition figures:

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
District	8	39,638	4,954.8
Condition within district	171	40,161	234.86

Assuming samples of equal size and equal variance within the several districts, that is, $s^2 = 234.86$, show that a sample of some 102 per district will be required to restrict the 5% fiducial limits to one unit (1%) condition on either side of the mean.

17.6—Sampling from two or more homogeneous populations. This problem has been considered in chapters 4, 6, 7, 9, and all following chapters. Estimates of the mean and variance of each sampled population were provided, together with tests of significance. In all of the foregoing we have assumed that the members of one population could be distinguished from those of another and that each population could be sampled independently. There remain some questions which we have not had occasion to consider heretofore.

Before proceeding, it may be wise to call attention to a prevalent fallacy. It is rather a common idea that always increased precision is attainable through larger sample size, since $s_2 = s/\sqrt{n}$. What is often overlooked is that this law applies only to sampling from a single homogeneous population. If the process of sampling carries you beyond the bounds of one population you may find that precision actually decreases as the sampling continues. It is to avoid this chance that experimental plots are grouped into blocks of uniform fertility, each block constituting a small population separately sampled. Similar devices are available in all circumstances where heterogeneity can be foreseen, as will be clear from the pages to follow. It is only necessary here to clarify thinking about the advantages of large samples. Since uniform experimental material is often limited in amount, several small samples, each from a homogeneous portion, usually yield more precise information than a single large one.

There is another fallacy about sample size sometimes observed in reports of sampling. An investigator may be confronted by a choice be-

tween securing a large number of interviews in a limited area where travel is cheap, or being reconciled to a smaller sample scattered at random over the entire region about which information is desired. The latter is the only way to insure equal chances of selection for every individual or sample in the population. Instances are on record where the entire sampling was confined to a selected area, such as a township, every individual in the chosen area being enumerated. While this may insure complete knowledge of the facts in the township, it concurrently insures complete ignorance of the facts in the remainder of the region. Representative samples must be scattered at random over the entire population even if only a small proportion of the individuals can be enumerated. A small sample properly designed contains information up to the limits of its size, but even a large sample confined to a portion of the population is devoid of information about the excluded portions.

In designing a sampling procedure involving a heterogeneous aggregate or *universe*, all available information can be used. Neyman has shown (6) that if the investigator knows the boundaries of the several homogeneous populations (subdivisions or strata), together with the size N and standard deviation, σ , of each, the most efficient sampling consists in taking at random k sampling units from each subdivision so that k is jointly proportional to N and σ . Suppose, for example, that corn yield per acre is normally distributed in each of three type-of-farming areas, as follows:

Area	Number of Acres of Corn N	Standard Deviation σ	Product $N\sigma$	Size of Sample $k = N\sigma(n/S)$
1	10,000	13	130,000	338
2	20,000	10	200,000	519
3	5,000	11	55,000	143
$T = 35,000$			$S = 385,000$	$n = 1,000$

The total sample, n , should then be divided into three parts proportional to the three products, $N\sigma$. To do this, add the three values of $N\sigma$, calling the sum, S . That is, $S = 385,000$. Decide upon the size of the entire sample, say $n = 1,000$. Multiply the quotient,

$$n/S = 1,000/385,000 = 0.0025974,$$

by each value of $N\sigma$ to get the corresponding k . For example, in area 1,

$$k = (130,000)(0.0025974) = 338$$

The three values of k should, of course, add to the chosen $n = 1,000$.

While exact knowledge of variation within the subdivisions is unusual, approximations of sufficient accuracy are often available. Previous

samplings or census data may be pertinent. A small experimental sample might be resorted to. Even general information about range may serve to approximate σ (sections 5.7 and 5.8).

One often has knowledge of the sizes of proposed subdivisions even if the standard deviations are unknown. If so, it is usually well to make the k 's proportional to the N 's. The investigation may not be so efficient as if the standard deviations were known, but will be as efficient as possible with the limited information available.

If there is no information about either N or σ , one should ordinarily make the sampling completely random, subdivision being inappropriate.

17.7—Analysis of data from two or more populations. We now suppose that the survey is completed and that the results are to be examined statistically. The methods are in the main familiar. For simplicity, let us consider first the case requiring no new devices at all, that in which the subdivisions are equal in size and in variance, only their means being different. The sample will then consist of equal numbers of items in the several strata. In table 17.1 is presented a summary of the

TABLE 17.1
ANALYSIS OF VARIANCE OF RESULTS OF A RANDOM SAMPLE OF 100 FROM EACH OF
THREE SUBDIVISIONS
Population of each subdivision, 10,000

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Total Subdivisions	299	14,755	
	2	1,600	800
Individuals	297	13,155	44.29

$\bar{x} = 71.21, s = 6.655$

results of such a sampling, the total sample of 300 equally divided among the three subdivisions.

Clearly the means of the subdivisions differed significantly so that the stratification is justified. In section 17.9 we shall evaluate the effect of the subdivision more precisely. Since the strata are known to be equal in size, the mean of the sample, 71.21, as well as the variance, 44.29, are unbiased estimates of the corresponding parameters of the universe of three populations. The variance of the mean is, therefore,

$$44.29/300 = 0.1476,$$

the standard error being 0.3842. For the 5% fiducial limits (section 3.8) we have $d.f. = 297$, $t_{.05} = 1.968$ and

$$\pm t_{.05} s_{\bar{x}} = \pm (1.968)(0.3842) = \pm 0.76$$

If we say, now, that the mean of the universe lies between $71.21 - 0.76 = 70.45$ and $71.21 + 0.76 = 71.97$, or that our sampling has led to a mean which is within $0.76/71.21 = 1.1\%$ of the true value, we shall be right unless a 1 in 20 chance has come off.

Ordinarily it is the universe total which must be estimated,

$$(30,000)(71.21) = 2,136,000,$$

with fiducial limits 30,000 times those of the mean, 2,114,000 and 2,159,000.

Having looked at the simplest problem in stratification, let us turn to the more realistic one in which the sizes of the subdivisions along with their means and their variances are all different. Results of a survey of this kind, such as might have been got from the procedure designed in the preceding section, are displayed in table 17.2. Numbers of acres in

TABLE 17.2
ANALYSIS OF VARIANCE AND SUMMARY OF THE DATA FROM A RANDOM SAMPLE OF
CORN YIELDS IN THREE TYPE-OF-FARMING AREAS

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Individual areas:			
1	337	55,689	165.25
2	518	54,742	105.68
3	142	17,044	120.03
Within areas	997	127,475	127.85
Type-of-farming areas	2	3,787.	1,893.50

Stratum number	Total Acres N	Sample Size k	Mean Yield (bu. per acre) $\bar{x}$
1	10,000	338	42.10
2	20,000	519	44.92
3	5,000	143	39.61
Total	35,000	1,000	

the three areas are assumed to be accurately known.

The estimated mean of the population is a weighted mean of the subdivision means, the weights being the subdivision sizes:

$$\text{weighted mean} = (SN\bar{x})/SN$$

Using the data of the example in hand,

$$\frac{(10,000)(42.10) + (20,000)(44.92) + (5,000)(39.61)}{10,000 + 20,000 + 5,000} = 43.35$$

bushels per acre. From the method of calculation, you see how important it is to have exact knowledge of the N 's, for even small errors in these weights may introduce noticeable bias in the estimate of the mean.

For ease in computation the weights may be reduced to proportional small numbers, such as 10, 20, and 5 (or 2, 4, and 1) in the example above; or to decimal fractions such as $2/7 = 0.28571$, $4/7 = 0.57143$, and $1/7 = 0.14286$, the sum of these decimal weights being unity. The latter is usually the easier method. Calculation then is confined to the sum of products, easily run on a machine:

$$(0.28571)(42.10) + (0.57143)(44.92) + (0.14286)(39.61) = 43.36$$

bushels per acre.

The estimated variance of the universe mean is also a weighted average of the variances of the several subdivision means, the weights being the squares of those for the mean:

$$\text{variance} = (SN^2s^2/k)/(SN)^2 = Sw^2s^2/k,$$

where w stands for *weight* and $Sw = 1$. In the numerical example,

$$(0.28571)^2 \frac{165.25}{338} + (0.57143)^2 \frac{105.68}{519} + (0.14286)^2 \frac{120.03}{143} = 0.12353$$

From the variance we calculate the standard error of the mean, 0.3515 bushels per acre.

Since the subdivision variances are not pooled in the manner of chapters 4 and 10, but are combined into a weighted average, one cannot assign degrees of freedom to the result. It follows that the ratio, $\bar{x}/s_x$, is not distributed as t , so that exact statements about fiducial limits cannot be made. Nevertheless, since the subsamples are large, the normal distribution ($t = \text{infinity}$) can be used with little lack of precision. With this qualification we may say that the 5% fiducial limits are

$$43.35 \pm (0.3515)(1.962),$$

that is, 42.66 and 44.04.

As before, the mean and fiducial limits, expressed as bushels per acre, may be expanded to production figures by multiplication:

$$\text{Total: } (35,000)(43.35) = 1,517,000 \text{ bushels}$$

$$\text{Limits: } 1,493,000 \text{ to } 1,541,000 \text{ bushels}$$

Let us pause to look at the formula (3) for the variance of the weighted mean of several subsamples drawn from populations of limited size (section 17.5),

$$S[N(N - k)s^2/k]/(SN)^2$$

If the samples comprise 5% or more of the populations, this formula should be substituted for the simpler one given above. If weights, w , are used in the calculations, and if $Sw = 1$, the formula just stated reduces to

$$S(w^2s^2/k) - (Sw s^2)/T,$$

where $T = SN$, the total for the universe. The second term constitutes a correction to be subtracted from the variance, $S(w^2s^2/k) = 0.12353$, already calculated. Since, in the corn yield data,

$$Sws^2 = (0.28571)(165.25) + \dots + (0.14286)(120.03) = 124.75,$$

the correction is $124.75/35,000 = 0.00356$. The resulting variance,

$$0.12353 - 0.00356 = 0.11994,$$

differs from the first in the third significant figure although the samples are only from 2% to 4% of the population sizes.

As is clear from the formulas which were used for weighted mean and its variance these estimates may be appreciably biased if the N 's are not accurately known. Gain in precision expected from stratification and large samples may be offset by poorly determined stratum sizes; in fact, serious loss of precision may occur. Cochran (3) gives a quantity by which the variance of the weighted mean is increased (as a first approximation) if there is some information as to the variance, s_N^2 , associated with the N 's:

$$S[(\bar{x} - \bar{x}_w)^2 s_N^2] / (SN)^2 = S[(\bar{x} - \bar{x}_w)^2 s_w^2]$$

where $\bar{x}_w$ is the weighted mean calculated above and s_w^2 is the variance of a weight.

If we assume for illustration that the w 's (or the N 's) are determined with coefficients of variability each 5%, we have the three standard deviations 0.014286, 0.028572, and 0.007143 acres with corresponding variances as set down in the table below. Also, $\bar{x}_w = 43.35$ bushels per acre. Substituting these values with the three stratum means, $\bar{x}$, found in table 17.2, the correction is calculated so:

Subdivision	$\bar{x} - \bar{x}_w$	$(\bar{x} - \bar{x}_w)^2$	s_w^2
1	-1.25	1.562	0.0002041
2	1.57	2.465	0.0008164
3	-3.74	13.988	0.0000510

Thus the correction to be added to 0.12353 is the sum of the products in the last two columns = 0.00304. This represents a decrease in information (section 10.12) of 2.4% attributable to errors in the weights. If the gain due to stratification were large (because of large differences, $\bar{x} - \bar{x}_w$) the loss from inaccurately determined weights would be correspondingly augmented.

Where knowledge of strata is vague, a well conducted sampling over the universe may be safer than attempted subdivision. How may that be handled? To illustrate the method, let us suppose that a random sample of 1,000 items has been drawn from the universe of type-of-farming areas which we have been studying. For purposes of comparing results, let us

assume that by a lucky coincidence the mean has turned out to be the same as the weighted mean of the three subsamples in table 17.2,

$$\frac{(338)(42.10) + (519)(44.92) + (143)(39.61)}{338 + 519 + 143} = 43.21 \text{ bushels per acre}$$

while the variance, by an equally happy chance, is, as before, 127.85. Since these are now taken to be the statistics computed from a random sample of 1,000, with no stratification, they are thereby unbiased estimates of the parameters. The standard error of the mean is

$$\sqrt{127.85/1,000} = 0.3575 \text{ bushels per acre}$$

To expand the sample mean to a production total for the universe, some estimate of the number of acres is required. Let us assume that this estimate is 35,000, but that now it is not accurately known. We have, then, the estimate of total production,

$$(35,000)(43.21) = 1,512,000 \text{ bushels}$$

If we are to arrive at an appropriate variance for this total we must first get some idea of the variance of the estimated size of the universe. This is likely not known with any precision. However, people familiar with the situation may venture the statement that 35,000 is probably within 1,000 of the correct number. If this is interpreted as meaning that the fiducial limits are $35,000 \pm 1,000$, a fair estimate of s , using $t = 2$, is $s = 500$. This reasoning is presented as one way of arriving at an approximate value of s . If more reliable information is to be had, use it.

If the squares of coefficients of variation are used in the calculation, the variance of a product of uncorrelated factors is, to a first approximation, the sum of their variances (7) (8). We, therefore, express our estimates of variance as follows:

$$\text{Yield per acre: } (100)^2(0.12786)/(43.21)^2 = 0.6848$$

$$\text{Acres: } (100)^2(500)^2/(35,000)^2 = 2.0408$$

The estimate of the variance of production in the three districts is, then,

$$0.6848 + 2.0408 = 2.7256,$$

the standard error being 1.6509%. Hence, the estimate of production has the standard error,

$$(1,512,000)(0.016509) = 24,960 \text{ bushels}$$

While an exact statement of fiducial probability would be inappropriate, we may set the analogous limits,

$$1,512,000 \pm 2(24,960),$$

that is, 1,462,000 and 1,562,000 bushels, and think of them as approximations—provided our estimate of the variance of population total was good.

The larger spread of these limits, compared to those set before, reflects the less accurate information which we are now assuming.

Those two percentage estimates of variance, 0.6848 and 2.0408, are informative. That of acre yield adds little to the total. Consequently, only slight precision can be gained for the production figure by refinements in sampling yield. If much narrower limits are to be set, it is the estimate of total acres that must be improved.

If an *attribute* is being sampled, and if it is assumed to be binomially distributed within the several strata, the means, p , and variances, pq/n , are substituted for $\bar{x}$ and s^2 in the formulas for estimating the mean and its standard error in the universe. As an example, consider a sample of 692 families in Iowa to determine, among other things, how many had victory gardens in 1943. The data were as follows:

Stratum	Number of Families	Weight, w	Number in Sample, n	Number With Gardens	Percentage With Gardens
Urban	312,393	0.4451	300	218	72.7
Rural	161,077	0.2297	155	147	94.8
Farm	228,354	0.3252	237	229	96.6
	701,824	1.0000	692	594	

The numbers of families were taken from the 1940 census. Since there was no information about variance, the sample was allotted approximately in proportion to the number of families per stratum, a sample of 1 per 1,000 being aimed at.

The weighted mean percentage of Iowa families having gardens was estimated as,

$$Spw = (72.7)(0.4451) + (94.8)(0.2297) + (96.6)(0.3252) = 85.5\%$$

The variance of this mean:

$$Sw^2pq/n = (0.4451)^2(72.7)(27.3)/300 + \text{etc.} = 1.62$$

The standard error, then, is 1.27%.

If approximately normal distribution is assumed in these large samples, one may set fiducial limits subject to the same reservations as before:

$$85.5 \pm (1.96)(1.27): 83.0\% \text{ and } 88.0\%$$

17.8—The structure of sampling investigations. In setting forth the randomized block and latin square designs of chapter 11, nothing was said about the possibility of sampling the plots instead of harvesting them completely. (For an example with references, see 9.) Each investigator, with little added expense, may plan his experiment so as to get information about the efficacy of this procedure. While we shall illustrate

EXAMPLE 17.5—The following estimates of cotton yield (pounds per acre) as of December 1, 1939, were reported by correspondents and furnished through courtesy of Samuel T. Marsh (5), Agricultural Statistician, U.S.D.A., Knoxville, Tennessee.

District	Weight, w	Number Reports, k	Degrees of Freedom	Sum of Squares	Mean Square, M	Mean Pounds Per Acre
1	0.359	77	76	566,614	7,455	343.7
2	0.466	184	183	908,046	4,962	233.2
3	0.076	40	39	128,145	3,286	188.2
4	0.060	31	30	169,994	5,666	266.4
5	0.016	5	4	31,330	7,833	287.0
6	0.023	9	8	89,800	11,225	265.0
	1.000	346	340	1,893,929	5,570.4	

Compute the weighted state mean, 273.03 pounds per acre, and its variance 20.527.

EXAMPLE 17.6—The analysis of variance of the Tennessee cotton yield estimates is this:

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Reports from same district	340	1,893,929	5,570.4
Districts	5	880,084	176,017
Tot. ¹	345	2,774,013	

Verify this analysis approximately by use of the data in example 17.5.

EXAMPLE 17.7—It is thought that the weights in the cotton data are not in error as much as 2%. Assume, then, that their standard deviations are one sixth of the ranges, $0.04w$. Show that this error in the weights decreases precision by less than 1%.

the method by use of a field plot layout for maize, all the principles apply directly to samplings of social or economic facts. The analogy will be pointed out later. (See also, 6 page 614.)

Our data are derived from a cooperative project of the United States Department of Agriculture and the Iowa Agricultural Experiment Station, the purpose being to study the development of the corn plant in relation to its environment. The layout was randomized blocks with four varieties replicated in five plots. Much of the growing material was cut during the summer for dry matter increase and chemical studies. Reserved for yield determination were eight *sampling units* in each plot. These units constitute the random elements in the sampling, their variation being the basis of *sampling error*. Shape, size and placement of the sampling units are a major problem in design. In the corn experiment, each consisted of four adjacent hills at the vertices of a rectangle. These hills, harvested separately, are the *units of measurement*, systematically placed in the sampling units.

Sampling units sometimes are not randomized over the whole plot but

are grouped into two or more *sampling areas* (subdivisions or strata). This precaution, often referred to as *local control*, insures more uniform coverage of the plot by the sampling units, preventing the possibility of their concentration in some one area. Variation in fertility over the plot thus tends to be equalized in the sampling. For the kind of study we are making there must be at least two sampling units in each sampling area. The entire structure of the experiment is set out in table 17.3.

TABLE 17.3
ANALYSIS OF VARIANCE OF SAMPLING DATA FROM A RANDOMIZED BLOCK FIELD
EXPERIMENT WITH MAIZE. AMES, IOWA, 1939. INDIVIDUAL HILL BASIS

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Variety	3	6.0723	
Block	4	1.7006	
Experimental error	12	1.6930	0.1411
Subdivisions within plots	20	2.9479	0.1474
Sampling units within subdivision	120	14.4387	0.1203
Hills within sampling units	480	38.9200	0.0811
Total	639	65.7725	

1 sampling unit		= 4 hills
1 subdivision	= 4 sampling units	= 16 hills
1 plot	= 2 subdivisions	= 32 hills
Entire experiment	= 20 plots	= 640 hills

The twenty whole plots constitute a randomized block experiment of the kind discussed in chapter 11. Here we are concerned not with tests of significance but rather with the composition of the *experimental error*—the possibilities of changing it by alterations in the design of the sampling.

A summary of the calculations may be convenient—they were explained more fully in section 10.9. After computation of the total sum of squares for the 640 hills, the steps are as follows, the same correction term being used throughout. (i) Compute the sum of squares for the 160 sampling units, each consisting of 4 hills, then deduct this from the total, leaving the sum of squares for hills within sampling units, 38.9200; (ii) compute the sum of squares for the 40 subdivisions, each containing 16 hills, and deduct this from the sum of squares for the 160 sampling units, leaving 14.4387, the sum of squares for sampling units within subdivisions; (iii) compute the sum of squares for the 20 plots, each having 32 hills, and subtract this from the sum of squares for the 40 subdivisions, the remainder, 2.9479, being designated as subdivisions within plots; (iv) the whole-plot analysis is then completed in the usual manner.

In table 17.3, the comparison that chiefly concerns us is that between the two mean squares representing random variation, the *sampling* and

experimental errors. The latter was examined rather carefully in section 11.3: it represents the variation of a set of deviations from regression values, assumed to be a randomly assorted lot from a normal population with mean, zero, and standard deviation, σ . Let us imagine the source of one of these deviations. If it is positive, the corresponding yield is more than was estimated for the plot, situated in its block and planted to its variety. Why? Perhaps in fertility it was particularly favored among plots of the same block. Perhaps also it was planted with seed that produced plants more vigorous than average. These advantages it might enjoy whether the harvesting was complete or by samples. But if only a sample is taken, the deviation may contain another component due to the random selection of an unusually productive batch of sampling units. Thus, each deviation from expected is made up of a portion peculiar to the plot and a second portion due to sampling within the plot. The variances of the two portions, being independent, are to be added.

In table 17.4 are copied the pertinent data about sampling and experimental errors. The variances of the means of sampling units and plots are

TABLE 17.4
SAMPLING AND EXPERIMENTAL ERRORS FROM TABLE 17.3, TOGETHER WITH
COMPONENTS OF VARIANCE

Source of Variation	Mean Square	Number of Units of Measurement	Variance of Means	Variance Is an Estimate of
Experimental Sampling	0.1411	32	0.004409	$\sigma_g^2/8 + \sigma_B^2$
	0.1203	4	0.030075	σ_B^2
		$s_g^2 = 0.030075$	$s_B^2 = 0.000650$	

calculated by dividing the corresponding mean squares by the respective numbers of hills (units of measurement). The sample estimates of the variance components are

$$s_g^2 = 0.030075$$

$$s_B^2 = 0.004409 - (0.030075)/8 = 0.000650$$

It should be observed that local control (that is, stratification) was quite effective in eliminating variation from experimental error. In all our discussion of this experiment, we are assuming that sampling will be confined to *these same strata*, so that differences between them do not affect our estimates of error.

The sample estimate of the variance of a plot mean may now be written,

$$s_z^2 = s_B^2 + s_g^2/k = 0.000650 + 0.030075/k,$$

where k is the number of sampling units per plot. For the experiment as designed, $k = 8$, and

$$s_z^2 = 0.000650 + 0.003759$$

To verify this, notice that the sum is 0.004409 as in the table.

It is now evident that the major portion of the variance of the plot mean has its source in the sampling of the plots. The implication is clear: to increase the precision of an experiment like this, alter the design so as to include more sampling units per half-plot (stratum). Suppose, for example, 24 sampling units were taken from each subdivision, then

$$\begin{aligned}s_B^2 &= 0.000650 + 0.030075/48 \\ &= 0.000650 + 0.000627 = 0.001277\end{aligned}$$

Thus, the *relative information* from this new design would be

$$\frac{1/0.001277}{1/0.004409} = \frac{0.004409}{0.001277} = 3.45$$

times that in the original, an increase of 245%. (See example 17.10 for some formulas involving "information.") But the increased information would be got at the expense of 6 times as much harvested crop. The efficiency of an experiment depends on the information per unit cost. Ordinarily, the entire crop is harvested anyhow, so that little extra cost is incurred from getting all the information. But in the experiment under consideration this was not possible because the other hills had been removed earlier for dry matter determinations.

For sake of illustration, let us now suppose that the variance of the hills within our sampling units is an unbiased estimate of the variance of hills selected at random. It probably is not, but is controlled in some manner by the systematic arrangement of the hills in the sampling units. Whether and how it would be affected by changing the pattern is a matter of conjecture. Nevertheless, by assuming that the hill variance would be unaltered if the size of the sampling unit were varied we can show how random data might be used for further study of the design. In table 17.5, the components of variance are estimated under this new hypothesis,

TABLE 17.5
COMPONENTS OF VARIANCE FOR TABLE 17.3 ON THE ASSUMPTION OF RANDOM
SELECTION OF THE HILLS

Source of Variation	Mean Square	Units of Measurement Hills	Variance of Means	Variance Is an Estimate of
Experimental error	0.1411	32	0.004409	$\sigma^2/32 + \sigma_B^2/8 + \sigma_B^2$
Sampling error	0.1203	4	0.030075	$\sigma^2/4 + \sigma_B^2$
Units	0.0811	1	0.0811	σ^2
	$\sigma^2 = 0.0811$	$\sigma_B^2 = 0.009800$	$\sigma_B^2 = 0.0006500$	

and we can write,

$$\begin{aligned}s_B^2 &= s_B^2 + s_B^2 k_B/k_B + s^2/k_B \\ &= 0.000650 + 0.009800/8 + 0.0811/32 \\ &= 0.000650 + 0.001225 + 0.002534,\end{aligned}$$

the sum being 0.004409 as before. The k 's are the numbers of hills per sampling unit and plot, respectively: k_s/k_p is the number of sampling units per plot.

As before, an increase in the numbers of hills and sampling units is indicated. If the number of hills were quadrupled and the number of sampling units doubled,

$$\begin{aligned}s_s^2 &= 0.000650 + 0.009800/16 + 0.0811/128 \\ &= 0.000650 + 0.000612 + 0.000634 = 0.001896,\end{aligned}$$

the relative information now being $0.004409/0.001896 = 2.33$ times the original amount, an increase of 133%, but at the expense of 4 times as many hills.

A more realistic question would be, "What would be the effect of decreasing the hills per sampling unit and increasing the number of sampling units proportionately, making no change in the amount of harvested material?" Take, for example, only 2 hills per sampling unit, allowing 8 sampling units per stratum:

$$\begin{aligned}s_s^2 &= 0.000650 + 0.00980/16 + 0.0811/32 \\ &= 0.000650 + 0.000612 + 0.002534 = 0.003796,\end{aligned}$$

an increase of 16% over the original information with no increase of expense except the inconvenience of locating the additional sampling units.

Is anything gained by sampling a plot rather than harvesting it entirely? This is the moot question as to sample *versus* census. Clearly, precision is at the maximum (sampling error = 0) if a census is taken. If there is no increase in cost of harvesting, in threshing losses, etc., then the complete harvest is obviously more efficient.

It should be noted that in the experiment being discussed there was little opportunity for changing the design—there was no further material available for sampling. Aside from the slight advantage of scattering the sampling units more widely (decreasing the number of hills in each) the only way to have gotten more information was to have increased the number of replications.

Let us look at another analysis of variance (table 17.6), not uncommon, in which experimental error is much greater than sampling error. Here s_s^2 is 0.01664 with $s^2/8$ only 0.0010. The reader may verify the fact that in these circumstances doubling the number of sampling units will increase information less than 3%, whereas an increase in replication from 5 blocks to 6 will make a corresponding increase of $1/5 = 20\%$. This emphasizes the fact that the sample-census problem should be examined anew in each situation where the answer is worthwhile. One object of this chapter is to set out some of the more useful methods.

In the social sciences, the sampling unit may comprise a family, the households along a city block, a township, a section of farm land, etc., the corresponding unit of measurement being perhaps the individual,

TABLE 17.6

ANALYSIS OF VARIANCE TO ILLUSTRATE SITUATION IN WHICH SAMPLING ERROR IS SMALL COMPARED TO EXPERIMENTAL ERROR. THE STRUCTURE OF THE EXPERIMENT IS THE SAME AS THAT OF TABLE 17.3

Source of Variation	Degrees of Freedom	Mean Square	Number Sampling Units, k	Components of Variance	Variance of Mean
Variety	3	2.0241			
Block	4	0.4252			
Experimental error	12	0.1411	8	$s_E^2 = 0.01664$	0.01764
Sampling error	120	0.0080	1	$s^2 = 0.00800$	

the family, the section, or the farm. It is the sampling unit that must be randomly selected. Its units of measurement form a fixed pattern. The entire population is subdivided (or stratified) into sampling areas like cities, income levels, or type-of-farming areas within which the population under investigation is supposed to be homogeneous.

Investigators who can spare the time and money to carry through the extra details of sampling may be well rewarded if they discover that inexpensive changes in design result in material increases in efficiency.

EXAMPLE 17.8—This is a modified form of the analysis of variance of the data collected in a wheat sampling survey (4) of 5 districts in North Dakota (1938):

Source of Variation	Degrees of Freedom	Mean Square	Number of Samples, k
Districts	4	329	88
Fields within district	215	65	2
Samples within field	220	19	1

Two samples were taken at random in each of 44 fields randomly selected in each of the 5 districts. Calculate the components of variance. Ans. $s_D^2 = 3$, $s_F^2 = 23$, $s^2 = 19$.

EXAMPLE 17.9—What is the relative information to be expected from each of the following arrangements? (i) double the sampling within each field; (ii) double the number of fields; (iii) take only one sample from each of double number of fields; and (iv) take 4 samples from each of half the number of fields. Ans. 1.03, 1.11, 1.08, and 0.88.

EXAMPLE 17.10—If s_0^2 is the original error variance and s^2 that resulting from some change in design, show by use of the definition in section 10.12 that:

$$(i) \text{ Relative information} = s_0^2/s^2$$

$$(ii) \text{ Increase in information} = (s_0^2/s^2) - 1$$

(i) may be greater than 1 or less, so that (ii) may be positive or negative. Both quantities may be expressed in percent by moving the decimal points 2 places to the right.

EXAMPLE 17.11—An experiment is being designed to test the nutritive value of 4 feed supplements. The weight of 5-week chicks is the variate to be measured. Ackerson and Muschl (1) reported the average weights and standard deviations of 5-week single comb white Leghorns:

Males: weight, 260 grams; standard deviation, 58 grams.

Females: weight, 242 grams; standard deviation, 49 grams.

It is sufficiently accurate to take the pooled bird variance as $s^2 = (58^2 + 49^2)/2 = 2,880$. The randomized blocks design is to be used with 6 replications and 10 chicks per cage, so that the analysis of variance is

Treatments	3	
Replications	5	
Error	15	$s^2_1 + \frac{1}{5}10s_B^2$
Chicks within cages	216	s^2_1

Assume the component due to environmental differences among the cages is 10% of 2,880 = 288 = s_B^2 . The error mean square will then be 5,760. What difference between two treatments would be significant at the 5% level? Ans. 30 grams. Calculate the coefficient of variation, assuming the average weight of all chicks to be 250 grams. Ans. 30%. If greater precision is desired, would you double the number of chicks per pen or increase the replications to 8? Re-read section 11.10; or better, sex the chicks and design the experiment accordingly.

17.9—Effectiveness of stratification. After examination of a stratified random sample, doubt may arise as to the advantage of the design or interest may be excited as to its efficiency. Especially if other samplings of the same universe are planned, it is desirable to learn if there is any advantage of the stratified sample over one taken completely at random.

An easy way to estimate the variance of a random sample from the results of a stratified sample is by constructing the analysis of variance of the sampled universe and from it calculating the total variance (mean square). The data of table 17.1 may be used to illustrate the case of equal populations with the same sized samples from them. Since $s^2 = 44.29$, the component for subdivision means is

$$s_M^2 = (800 - 44.29)/100 = 7.5571$$

With these components, an analysis of variance for the universe is built up in table 17.7. The mean square for subdivisions is

$$s^2 + Ns_M^2 = 44.29 + (10,000)(7.5571) = 75,615,$$

TABLE 17.7
ANALYSIS OF VARIANCE OF UNIVERSE CONSTRUCTED FROM ESTIMATES OF σ^2 AND σ_B^2
Data from Table 17.1

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Units within strata	29,997	1,328,600	44.29
Strata (or subdivisions)	2	151,200	75,615.
Totals	29,999	1,479,800	49.33

$N = 10,000$ being the size of each population. The first two mean squares in the table, multiplied by the corresponding degrees of freedom for the universe, yield sums of squares that add to the total. Finally, the mean square, 49.33, is the desired estimate of the variance in the unstratified universe.

It can now be said: (i) That the information in the stratified sample is

$$49.33/44.29 - 1 = 0.114 \text{ or } 11.4\%$$

greater than that in the random; or (ii) that the information lost by abandonment of stratification would be

$$1 - 44.29/49.33 = 0.102 \text{ or } 10.2\%$$

Since the first of the foregoing statements is the one usually interesting to samplers, a formula may be more convenient than construction of the analysis of variance. The gain in information due to stratification is

$$\frac{N(m-1)s_M^2}{(Nm-1)s^2},$$

m being the number of strata. Substituting, this gain in our example is

$$\frac{(10,000)(3-1)(7.5571)}{(30,000-1)(44.29)} = 11.4\%, \text{ as above.}$$

Since Nm is often large, as here, it is then adequate to write the formula,

$$\frac{N(m-1)s_M^2}{Nms^2} = \frac{m-1}{m} \cdot \frac{s_M^2}{s^2}$$

The present data entered in this formula give the same result as before. These formulas apply only to strata of equal size.

The solution of the problem with strata of varying size parallels that just given. The data in table 17.2 will be used as illustration.

To calculate s_M^2 , the formula for k_0 (section 10.8) is required:

$$k_0 = \frac{1}{m-1} \left(Sk - \frac{Sk^2}{Sk} \right) = \frac{1}{2} \left(1,000 + \frac{404,054}{1,000} \right) = 297.97$$

From this,

$$s_M^2 = (1,893.50 - 127.85)/297.97 = 5.9256$$

We shall require, also, an average of the stratum sizes, for which the same as above formula is used:

$$N_0 = \frac{1}{m-1} \left(SN - \frac{SN^2}{SN} \right) = \frac{1}{2} \left(35,000 - \frac{525,000,000}{35,000} \right) = 10,000$$

Next, the analysis of variance in table 17.8 is constructed. For the sums of squares in the several strata, the mean squares of table 17.2 are multiplied by the degrees of freedom for the populations. The average mean square within the strata is the weighted mean,

$$s_P^2 = 4,365,800/34,997 = 124.75$$

The mean square for the stratum means is calculated on the assumption that it has the components, $s_M^2 = 5.9256$ and $s_P^2 = 124.75$; then

$$s_P^2 + N_0 s_M^2 = 124.75 + (10,000)(5.9256) = 59,381$$

TABLE 17.8
ANALYSIS OF VARIANCE OF UNIVERSE CONSTRUCTED FROM ESTIMATES IN TABLE 17.2

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Individual strata			
1	9,999	1,652,300	165.25
2	19,999	2,113,500	105.68
3	4,999	600,000	120.03
Within strata	34,997	4,365,800	124.75 = s_p^2
Strata	2	118,800	59,381
Total	34,999	4,484,600	128.14

is the mean square to be multiplied by $d.f. = 2$ to get the stratum sum of squares, 118,800.

Finally, the total sum of squares divided by $d.f. = 34,999$ gives the mean square, 128.14, an estimate of the variance of a sample randomized over the universe, stratification ignored.

We are now able to calculate the variance of the mean of the random sample, $128.14/1,000 = 0.12814$, and to compare it with that of the stratified sample, 0.12353 (section 17.7). The increase in information due to stratification is

$$(0.12814/0.12353) - 1 = 0.037 \text{ or } 3.7\%$$

Two conclusions may be drawn. (i) Although the differences among strata are highly significant (table 17.2) the advantage to be gained by stratification of this sample is scarcely worthwhile if it costs much. That is because the differences among the stratum means are slight. (ii) Inaccuracies in the weights might easily nullify the gain from stratification.

The construction of the analysis of variance may be avoided by use of the following formula for the information in a stratified sample relative to that in the random:

$$\frac{[1 + 1/(T - m)][SwM - SM/T] + (1 - Sw^2)s_M^2}{(1 - 1/T)(Sk)(Sw^2M/k)}$$

$T = SN = 35,000$, the sum of the sizes of the m strata; w refers to the weights of the strata (0.28571, 0.57143, and 0.14286 of section 17.7); M stands for the mean squares of the several strata (165.25, 105.68, and 120.03 of table 17.2); $Sk = 1,000$, the total sample size; and $Sw^2M/k = 0.12353$, the variance of the mean of the stratified sample. Substitution of the various quantities reduces the relative information to 1.037, the increase due to stratification being 3.7% as before.

If T is large and m small the factors, $[1 + 1/(T - m)]$ and $1 - 1/T$, are each so close to unity as to be negligible.

EXAMPLE 17.12—Show that N_0 may be computed from the formula,

$$N_0 = T(1 - Sw^2)/(m - 1)$$

In the Tennessee cotton data of example 17.5, T was 707,000 acres. Calculate $N_0 = 91,026$ acres.

EXAMPLE 17.13—From the analysis of the sample of cotton yields of example 17.6, calculate $k_0 = 44.662$ and $s_M^2 = 3,816.4$.

EXAMPLE 17.14—Compute the increase in information due to stratification in the cotton data, noticing $Sw^2M/k = 20.527$ in example 17.5. Ans. 18.5%.

EXAMPLE 17.15—Show that the weighted mean of the population (strata) mean squares is given by,

$$s_P^2 = \frac{TSwM - SM}{T - m},$$

which reduces to SwM if T is large compared to m and SM . For the cotton, $s_P^2 = 5,961.78$

EXAMPLE 17.16—Derive the formula for relative information by following the steps described for Table 17.8.

EXAMPLE 17.17—Using the weights and variances of the 6 districts in example 17.5, design a sampling in which the k 's are proportional to $N\sigma$ and $Sk = 346$. Ans. $k = 140, 149, 20, 20, 6$, and 11.

REFERENCES

1. C. W. ACKERSON and F. E. MUSSEHL. *Journal of Agricultural Research*, 40:863 (1930).
2. W. G. COCHRAN and D. J. WATSON. *Empire Journal of Experimental Agriculture*, 4:69 (1936).
3. W. G. COCHRAN. *Journal of the American Statistical Association*, 34:492 (1938).
4. ARNOLD J. KING and EMIL H. JEBB. *Iowa Agricultural Experiment Station Research Bulletin No. 273* (1940).
5. SAMUEL T. MARSH. Personal communication (1940).
6. JERZY NEYMAN. *Journal of the Royal Statistical Society*, 97:558 (1934).
7. KARL PEARSON. *Proceedings of the Royal Society of London*, 60:489 (1897).
8. LOWELL J. REED. *Journal of the Washington Academy of Science*, 11:449 (1921).
9. F. YATES and I. ZACOPANAY. *The Journal of Agricultural Science*, 25:545 (1935).
10. F. YATES. *Annals of Eugenics*, 6:202 (1935).
11. G. UDRY YULE and M. G. KENDALL. *An introduction to the Theory of Statistics*, 11th Edition. Charles Griffin and Co., Ltd., London.

INDEX OF MATHEMATICAL TABLES

Arc $\sin \sqrt{\text{percentage}}$	449
Chi-square	190
Coefficients and divisors for orthogonal comparisons in regression	410
Coefficients and polynomials for terminal values and differences	393
Confidence interval, 5% and 1%	4
F , the variance ratio	222
Normal frequency curve, ordinates	179
Normal frequency curve, cumulative frequency (probability)	180
R , the multiple correlation coefficient, 5% and 1% points	351
r , the correlation coefficient, 5% and 1% levels	149
r_s , Spearman's rank correlation, 5% and 1% levels	166
Random digits	10
Range/ σ , mean values, 5% and 1% points	98
t , 5% and 1% levels	65
z and r , graph	152

GENERAL INDEX AND INDEX OF SYMBOLS

Absolute value, 36, 39 Ackerson, C. W., 472, 476 Addition theorem chi-square, 189 sum of square and degrees of freedom, 127, 216, 255 Adjusted value of chi-square, 22, 26, 193, 197, 199 mean, 116, 319, 321, 330 validity of use, 334, 338 R, 348 Y, 114 Allan, F. E., 268, 316 American Medical Association, 31, 53 Analysis of variance group comparisons, 214 individual comparisons, 253 in regression, 127, 157, 273, 348 of subsample means, 236, 273, 279, 301, 400 Anderson, R. D., 410, 430 André, Floyd, 209, 213, 378, 382, 399, 440, 452 Angle transformation, 447 Arbitrary origin, G, 89, 170 Arc sine transformation, 447 Arkin, Herbert, 20, 30 Array, 32, 55 of means, 408 Asterisk *significant, 226 **highly significant, 226, 266 Asymmetry; <i>see</i> Distribution, asymmetrical Attributes, 1, 188, 431 presence recorded quantitatively, 210 sampling of, 7	Autocatalytic growth, 399 Average of concentration or central tendency, 2, 32, 50, 57, 143 of ratios, rates, and percentages, 27 of relation correlation coefficient, 151, 243, 357 regression coefficient, 105, 108, 340 of variation, 31, 36, 47, 49, 50, 60, 61 Ayres, J. C., 43, 53 b, regression coefficient, 106 $b_{Y1.2}$, partial regression coefficient, 340, 364 $b'_{Y1.2}$, standard partial regression coefficient, 342 Baker, C. T., 421, 430 Baker, G. A., 148, 163, 167 Baker, W. G., 69 Barlow, Peter, 175, 186 Barnard, M. M., 354, 373 Bartlett, M. S., 200, 213, 249, 252, 339, 412, 419, 430, 446, 451 Batchelor, L. D., 313, 316 Bateson, W., 193, 213 Beadles, Jesse R., 45, 53 Beale, Helen Purdy, 44, 53 Becker, E. R., 293, 316 Behrens, W. U., 83, 88 Bell, Martin A., 27, 30 Bias, 2, 22, 106, 193, 244, 456 in r or z, 154, 244 Bigger, J. H., 125, 137 Binomial coefficients, 433 Binomial distribution; <i>see also</i> Distribution large samples, 431 small samples, 4, 188
---	--

- Binomial theorem, 431
- Bivariate population, 138
- Bliss, C. I., 399, 446, 449, 451
- Blocks, randomized, 253, 265, 266, 412
 - covariance in, 331, 354
- Boatman, J. L., 69
- Boyd, Mark F., 199, 213
- Boynton, Bernice, 40, 48, 53
- Brady, J., 411, 430
- Brandt, A. E., 46, 53, 122, 137, 141, 167, 206, 281, 316, 421, 422, 430
- Breneman, W. R., 75, 85, 88, 111, 137
- Brown, Bernice, 289, 316
- Brown, P. E., 82, 88
- Brunson, Arthur M., 358, 373
- Bryan, A. A., 54, 74
- Burnett, L. C., 208, 213
- Burroughs, Wise, 45, 53

- C, correction term, 219
- Caffrey, D. J., 200, 213
- Calculating machine, 91, 121
- Cannon, C. Y., 423, 430
- Case study, 114
- Catchpole, H. R., 82, 88
- Central tendency; *see* Average
- Chi-square, χ^2 , 16, 250, 315, 435, 438, 443
 - addition theorem, 189
 - adjusted value of, 22, 26, 193
 - frequency distribution of, 18, 190, 194
 - heterogeneity, 191
 - index of dispersion, 16
 - interaction, 191, 212
 - calculated from percentages, 25
 - and sample size, 24, 26
 - sampling of, 17
 - significance of, 21, 182
 - small expected numbers, 22, 193
 - special formulas, 16, 26, 196, 198, 202, 206, 208, 210, 211, 250
 - table of, 21, 190
- Clarke, G. L., 304, 317, 451, 452
- Class bounds, 55
- Class frequency, f , 18, 55, 159
- Class interval, I , 18, 171
 - unequal, 20, 63
- Class mark, 55, 169
- Clements, Frederic E., 52, 53
- Clopper, C. J., 2, 30
- Cochran, W. G., 23, 30, 83, 88, 268, 301, 305, 316, 412, 414, 430, 446, 451, 455, 458, 463, 464, 476
- Code numbers, 92, 93, 123, 141, 169, 217
 - rules and precautions about, 94
- Coefficient
 - binomial, 433
 - correlation, 138
 - multiple, 347
 - partial, 357
 - regression, 106
 - partial, 340
 - standard partial, 342
- Coefficient—*Continued*
 - of variation, C , 40, 46, 47, 80, 266, 273, 280, 457
- Coefficients for sets of independent comparisons, 401, 410
- Collins, E. V., 124, 137
- Collins, G. N., 69, 74
- Colton, Raymond R., 20, 30
- Combinations, 431
- Common causes, 162
- Common elements, 145, 157
- Common sense in computation, 99
- Comparisons
 - based on experimental results, 408
 - group, 75, 80
 - in analysis of variance, 214
 - in regression, 114
 - individual, 31, 43
 - in analysis of variance, 253
 - in regression, 114
 - individual *vs.* group, 84
 - nonorthogonal sets of, 406
 - orthogonal sets of, 403, 412
- Components of error, 413
- Components of variance; *see* Variance
- Compound interest, law of, 374
- Confidence inference, 3, 41
- Confidence interval, 3, 14, 435
 - table of, 4
- Confounding, 79
- Constants, fitting, 296
- Continuous variate or variable, 31, 171
- Control
 - local, 468
 - statistical, 126
- Correction
 - for continuity in chi-square, 22, 193
 - for mean, 89
 - for sum of squares, 89
 - Sheppard's, 173
- Correlation, 138
 - average of, 150
 - coefficient, r , 138, 139
 - variation in sampling, 145
 - and differences, 159
 - estimate of, 148
 - interclass, 138, and intraclass, 243
 - in large samples, 183, 185
 - between mean and variance, 315, 445
 - multiple, R , 347, 368
 - multiple, adjusted, R_A , 348
 - nonsense, 164
 - organic, 141, 164
 - partial, $r_{12.3}$ or $r_{12.4}$, 357
 - product moment, 138
 - of ranks, r_s , 164
 - with ratios, 162
 - and regression, 142, 155, 157, 185
 - spurious, 162
 - with sums, 162
 - table of 5% and 1% levels, 149
 - total or zero order, 138

- Council on foods, 32, 53
 Covariance, 108, 144, 318
 departure from linearity in, 384
 in groups, 325
 in latin square, 337
 multiple, 350
 in groups, 350
 in randomized blocks, 331
 two criteria of classification, 354
 Cox, Gertrude M., 83, 88, 268, 295, 316, 317, 339
 Crampton, Earl Wilcox, 46, 53, 98, 102, 134, 137, 336, 338
 Crathorne, A. R., 141, 168
 Criteria of classification
 one, 214, 318
 two, 253, 331, 354
 with two or more items in sub-classes, 275
 three or more, 304
 Culbertson, C. C., 78, 88, 141, 168, 262, 317, 324, 338
 Cumulative percentage distribution of t , 64
 Curve fitting, 374
 Curvilinear regression, 374
- $d.f.$; *see* Degrees of freedom
 $d_{y..}$ or $d_{y..}$, deviation from regression, 106, 364
 Darwin, Charles Robert, 70, 74
 Davenport, C. B., 175, 186
 David, F. W., 155, 168
 David, Pedro A., 62, 74
 Davidson, O. W., 93, 102
 Davis, Glen N., 200, 212, 213
 Dawes, Ben, 131, 137
 Dawson, W. T., 382
 Dean, Harold L., 78, 88
 Decker, George C., 209, 213, 378, 382, 399, 440, 452
 Degrees of freedom, $d.f.$, 37, 434
 individual or single, 43, 131, 191, 227, 263, 267, 273, 303, 330, 400
 Degrees of freedom in
 binomial distribution, 189, 437, 438
 components of error, 279
 correlation, 149
 covariance, 320
 enumeration data, 189, 192, 204, 437
 group comparisons, 76, 215
 individual comparisons, 44, 400
 multiple regression, 348
 partial correlation, 358
 Poisson distribution, 443
 regression, 119, 131
 Deletion of a variate, 369
 Deming, W. Edwards, 105, 137, 374, 399
 Design of
 experiments; *see* Experiment
 samplings, 453
 Deviate, standard, x' or t , 144, 343
- Deviation
 absolute, 36
 from expected, 16
 from mean, x , 34
 probable, $p.d.$, 49
 from regression, $d_{y..}$ or $d_{y..}$, 107, 113, 364, 368, 381, 382, 384
 standard; *see* Standard deviation
 relative; *see* Coefficient of variation
 Difference between means, 75
 Differences, 66
 and correlation, 159
 variance of 67, 76, 160, 181
 Digits, significant, 95
 Direct method in multiple regression, 364
 Discontinuous variable or variate, 21, 54, 193, 440
 Discrepancy, 255
 second order or three factor, 306
 Discrete variate or variable, 21, 440
 Disproportionate subclass numbers, 284
 Distribution, 14
 asymmetrical, 20, 60, 148, 174
 binomial, 188, 431, 447, 458, 466
 of chi-square, 21, 189
 of correlation coefficient, 147
 cumulative percentage, 64
 of F , 219, 222
 frequency, 18, 55, 169, 173
 geographical, 54
 normal; *see* Normal distribution
 Poisson, 188, 210, 440, 446
 of range/ σ , 97
 of sample means, 57
 of standard deviation, 60
 of t , 15, 62
 of variance, 60
 of z , 150
 Dobzhansky, Th., 203, 213
 Donaldson, Henry H., 48, 53
 Doolittle, M. H., 359, 373
 Dosage-mortality curve, 399
 Dungan, George H., 108, 125, 137
 Durost, N., 19, 30
 Dwyer, Paul S., 359, 373
- e , base of Napierian logarithms, 376, 441
 Eden, T., 148, 168
 Edwards, T. I., 100, 102, 171, 186
 Effects, 426, 428
 Efficiency of an experiment, 72, 246, 264, 301, 470, 473
 Ekas, Merle P., 175, 186
 Elements, common, 145, 157
 Ellipse, 140
 degenerate, 141
 Emmert, E. M., 388, 399
 Engledow, F. L., 214, 252
 Enumeration data, 7, 188, 315, 431
 Error, 106
 components of, 413
 estimate of, 43, 79, 118, 275, 313, 420

Error—Continued

- of estimate, 107, 328, 346, 381
 - standard,
 - $s_{y \cdot x}$, 118
 - $s_{Y \cdot X}$, 347
 - experimental, 43, 75, 320, 357, 468
 - probable, 49
 - regression, 323, 357
 - standard, s_e , 43, 61
- Espe, Dwight, 294, 316
- Estimate, 60, 149, 346, 374
 - unbiased, 2, 8, 14, 31, 38, 39, 58, 60, 73, 118, 255, 286
 - of variance, 38, 67, 159, 214, 228, 255, 259, 260, 261, 272, 278, 307
- Estimated value in regression, $\hat{Y}$, 107
- Evidence, 15, 19, 21, 23, 47, 195, 453
- Evvard, John M., 141, 168
- Expected numbers or frequencies; *see* Frequencies
- Expected values
 - in latin square, 274
 - in randomized blocks, 257
- Experiment
 - complex, 301, 313
 - design of, 267, 273
 - efficiency of, 72, 264, 301, 473
 - factorial, 404, 423
 - randomized blocks, 412
 - reversal or switch-back, 421
 - split plot, 309
 - and statistics, 87
- Experimental error; *see* Error
- Experimental technique, 79, 88
- Exponential growth curve, 374
- Extrapolation, 382
- Eye-fitted curves, 398
- Ezekiel, Mordecai, 359, 373
- F , variance ratio, 218, 231
- F table, 222
- $F = t^2$, 227
- f , frequency, 169
- Factorial experiment, 404, 423
- Fiducial inference, 41, 66, 73
- Fiducial interval, 42, 64
- Fiducial limits, l , 42, 456
 - in correlation, 148
 - in multiple regression, 366, 368
 - in regression, 119
- Fiducial statement, 64, 118, 148, 382
- Field plan, 254, 265, 272
- Finite differences, 374
- Fischer, Carl H., 145, 157, 168
- Fisher, R. A., 16, 23, 30, 42, 53, 65, 70, 74, 83, 88, 130, 137, 148, 150, 154, 168, 174, 186, 191, 194, 207, 213, 214, 219, 252, 266, 268, 271, 316, 339, 369, 373, 388, 399, 403, 410, 423, 430

- Fitting constants, 296
- Flint, L. H., 69, 74
- Flint, W. P., 125, 137
- Forecast; *see* Estimated value in regression
- Forster, H. C., 266, 316
- Fourfold table, 194
- Freedom, degrees of; *see* Degrees of freedom
- Freeman, Frank N., 244, 252
- Frequencies, expected, 16, 188, 194, 432
- Frequency, f , 169
 - relative, 431
- Frequency curve, 178
- Frequency distribution (*see also* Normal frequency), 18, 55, 169, 173
 - cumulative, 21, 64
 - of chi-square, 18, 21, 190
- Fryer, H. C., 23, 30
- G , arbitrary origin for X , 89
- g_1 , a measure of skewness, 174
- g_2 , a measure of kurtosis, 176
- Gaines, J. C., 447
- Galton, Francis, 70, 74, 103, 132, 137
- Gauss, Carl Friedrich, 366, 373
- Geddes, W. F., 378, 399
- Geographical distribution, 54
- Geometric mean; *see* Mean
- Gerdell, R. W., 146, 168
- Goodness of fit, 396
- Goulden, C. H., 23, 30, 194, 213, 425, 430
- Gowen, J. W., 23, 30, 200, 213, 287, 293, 297, 316, 377
- Grant, Joseph D., 374, 399
- Graphical representation, 19, 33
- Gray, Horace, 43, 53
- Gray, P. H. H., 210, 213
- Group comparisons
 - difficulties met, 79
 - in regression, 114, 318, 350
 - two groups, 75
 - more than two groups, 214
- Groups with different numbers of individuals, 80, 232
- Grout, Roy A., 142, 168, 359, 373
- Grove, L. C., 45, 53
- Growth curves, 103
 - exponential, 374
- Growth rate, relative, 101
- Haber, E. S., 147, 162, 168, 415, 430
- Hall, A. D., 54, 74
- Hall, Phoebe R., 293, 316
- Hammond, W. E., 78, 88
- Hansberry, T. Roy, 113, 116, 137, 200, 213, 227, 252
- Hansen, E. N., 423, 430
- Harris, J. A., 223, 246, 252
- Harrison, Carter M., 94, 102
- Harshaw, H. M., 339
- Haskell, A. C., 337, 399
- Hayes, H. K., 305, 316

- Heterogeneity chi-square, 191
 Hierarchical subdivisions, 207
 Histogram, 20, 55
 Hoffman, Katherine L., 261, 316
 Holbert, J. R., 125, 137
 Holland, R. H., 398, 399
 Holzinger, Karl J., 244, 252
 Homogeneity, 191, 212, 437
 of variance, 249
 Homogeneous population, 454
 Homogeneous variance, 105, 107, 108
 Hotelling, Harold, 147, 165, 168
 Houseman, E. E., 410, 430
 Hoyt, Elizabeth E., 93, 102
 Hypothesis
 of homogeneity, 212
 of linear regression, 381, 382
 9:3:3:1, 192
 null, 43, 47, 66, 82, 118
 about population, 15
 test of, 44, 66, 73, 118, 148, 218, 238, 259, 278, 303, 308, 314, 382
 3:1, 24, 188

I, class interval, 173
 Immer, F. R., 305, 316
 Independence, test of, 194, 204, 212
 Independent variate, X , 105
 Individual comparisons, 43, 72, 84, 160, 400
 in regression, 114, 409
 Individual degrees of freedom, 400
 Individual *vs.* group comparisons, 84
 Inferences about population, 2, 31, 58, 64, 118, 148, 453
 Information, 72, 248, 266, 470, 472, 473
 Interaction, 191, 201, 204, 212, 277, 285, 287, 405, 425, 428
 second order or three factor, 202, 308, 426, 428,
 Interclass correlation; *see* Correlation
 Interpolation
 curved regression, 374, 398
 linear regression, 107
 Interquartile interval, $Q_3 - Q_1$, 51
 Interval
 class, I , 18, 171
 confidence, 3
 fiducial, I , 42, 64, 119, 150, 369
 interquartile, 51
 Intraclass correlation, r_t , 243
 Inverse operations, 93
 Irrational number, 441
 Irwin, M. R., 206, 213

 Jebe, Emil H., 457, 472, 476
 Jenkins, M. F., 27, 30, 125, 137, 235, 252
 Jessen, Raymond J., 204, 213, 291, 295, 317

k, number of items in a group or sub-sample, 214

k statistics, 174
*k*₀, average number of items, 234
 Kendall, M. G., 165, 168, 454, 476
 Kendall, S. F. H., 165, 168
 Kenney, John F., 74
 Kiesselbach, T. A., 99, 102
 King, Arnold J., 457, 472, 476
 Kirk, Esben, 52, 53
 Kitchen, S. F., 199, 213
 Kurtosis, 176

 Lambert, W. V., 27, 30
 Large sample methods, 169
 Latin square, 271
 Least count, 54
 Least significant mean difference; *see* Significant difference
 Least squares, 39, 111, 340
 Lee, Alice, 133, 137, 138, 168
 Leggatt, C. W., 441, 444, 452
 Lewis, William H., Jr., 52, 53
 Li, H. W., 271, 316, 336, 338
 Limits, fiducial; *see* Fiducial limits
 Lindstrom, E. W., 172, 186, 188, 208, 213
 Linear regression, 103, 340, 346
 Lipka, Joseph, 374, 399
 Liu, T. N., 271, 316, 336, 338
 Logarithmic scale, 375
 Logarithmic transformation, 448
 Logarithm
 common, 249, 304, 377, 448
 Napierian or natural, 376
 Logistic, 399
 Long, Frances C., 52, 53
 Love, H. H., 46, 53
 Lowe, Belle, 216, 252, 261, 316
 Lush, Jay L., 151, 168

m
 expected number, 16, 34
 mean of population, 38
 MacArthur, John W., 24, 30
 Machine calculation, 92, 121, 290
 Magistad, O. C., 49, 53
 Mahalanobis, P. C., 219, 252
 Marsh, Samuel T., 467, 476
 Mather, K., 191, 207, 213
 "Mathetes," 210, 213
 Mayr, Ernst, 203, 213
 McLane, J. W., 69, 74
 McLaughlin, Laura, 93, 102
 McPhee, Hugh C., 436, 452
 Mean
 adjusted, 319, 321
 arithmetic, $\bar{x}$, 31, 169
 geometric, 143, 144, 358, 451
 harmonic, 288
 Mean absolute deviation, 36
 Mean difference, 43
 Means
 method of unweighted, 287, 294, 299
 method of weighted squares, 291, 299

- Mean square, 37, 215, 259
 short method of calculation, 400, 410
 Measurement data, 7, 31
 Median, Q_2 , 50, 220
 Meng, C. J., 271, 316, 336, 338
 Mercer, W. B., 54, 74
 Method of averages, 105, 107
 Metzger, W. H., 379, 399
 Missing plot or data, 268, 274
 Mitchell, H. H., 45, 53
 Mitscherlich equation, 378
 Morgan, Ethel Cessna, 93, 102
 Morris, V. H., 146, 168
 Mortality curve, dosage, time, 399
 Muench, Hugo, 199, 213
 Multiple
 correlation coefficient, R , 347
 table of 5% and 1% points, 351
 covariance
 in groups, 350
 two criteria of classification, 354
 regression, 340
 Multipliers, Gauss, 366, 367
 Musschl, F. E., 472, 476

 N , number of individuals in population, 460
 n , number of
 groups or subsamples, 214
 items in sample, 34
 Napierian or natural logarithm; *see* Logarithm
 Nelson, P. Mabel, 93, 102
 Newman, Horatio H., 244, 252
 Neyman, J., 460, 476
 Nonorthogonal analysis, 285
 Nonorthogonal sets of comparisons, 406
 Normal distribution, 37, 42, 54, 169, 221, 259
 table of, 180
 test of, 174, 176
 Normal equations, 342, 361, 364, 367
 Normal frequency curve, 177
 equation of, 183
 table of ordinates, 179
 Normally distributed sample, 55
 Null hypothesis, 43, 47, 66, 82, 118, 278
 Numbers; *see* Expected

 O'Neal, James R., 423, 430
 Operations, inverse, 93
 Origin, arbitrary, G , 89
 Origin of deviations, 110
 Orthogonal
 polynomials, 388, 409
 comparisons, sets of, 403, 412
 Outcome, 43, 85, 253, 260, 262, 304

 P or p , probability, 77, 432
 $p.d.$, probable deviation, 49
 Pabst, Margaret Richards, 147, 165, 168
 Page, Irvine H., 52, 53

 Paired observations; *see* Individual comparisons
 Parabola, 379, 392, 394
 Parameter, 38
 Park, O. W., 78, 88
 Parker, E. R., 313, 316
 Parkes, A. S., 437, 452
 Partial correlation, $r_{12.3}$, 357
 Partial regression coefficient; *see* Regression coefficient
 Pascal's triangle, 433
 Pearl, Raymond, 48, 53, 365, 373, 399
 Pearson, E. S., 2, 30, 98, 102, 221, 252, 339
 Pearson, Karl, 16, 30, 133, 137, 162, 168, 444, 452, 465, 476
 Pearson, P. B., 82, 88
 Penquite, Robert, 374, 399
 Percentage, 27, 315, 447
 chi-square from, 25
 Perennial crop, 415
 Permutations, 431
 Plants, perennial, 313
 Plot, missing, 268, 274
 Poisson distribution
 large samples, 446
 small samples, 188, 210
 Polynomial
 orthogonal, 388, 409
 second degree, 379, 394
 Pooled sum of squares, 77, 215, 326
 Pooled variance, 76

 Population, 2, 29, 38, 44, 54, 76, 79, 82, 104, 106, 119, 148, 188, 214, 217, 228, 259, 303, 349, 434, 453
 small, 458
 Porter, R. H., 269, 316
 Powers, LeRoy, 305, 316
 Predicted value of Y , $\hat{Y}$, 107, 257, 346, 374
 Price, W. C., 377
 Principle of least squares, 39
 Probability, p or P , 3, 7, 22, 42, 63, 98, 180, 189, 220, 350, 432
 Probable deviation, $p.d.$, 49
 Product moment correlation; *see* Correlation
 Proportional subclass numbers, 281
 Punched cards, 122
 Punnett, R. C., 193, 213

 Q , quartiles, 50
 q , probability that event will not occur, 432
 Quartiles, Q_1 and Q_3 , 50

 R , coefficient of multiple correlation, 347
 table of 5% and 1% points, 351
 R_A , adjusted R , 348
 r , correlation coefficient, 138
 table of 5% and 1% levels, 149
 $r_{12.3}$ or $r_{12.4}$, partial correlation, 357

- ρ , correlation in population, 148
 $R \times C$ table, 204, 293
 $R \times 2$ table, 205, 289
 Random differences, 67, 73, 159
 Random numbers, table of, 10
 Random sample; *see* Sample
 Randomized blocks; *see* Blocks
 Randomness, 8, 267, 275
 Range, 31, 40, 70, 451
 of code numbers, 93
 of means, 58
 significance of, 97, 409
 Range/ σ , 97, 409
 Rank correlation, r_s , 164
 Rates, 27, 134
 and percentages, 100
 Ratio paper, 375
 Ratios, 27, 134
 Rectification, 375
 Reddy, Chas. S., 35, 53
 Reed, H. S., 398, 399
 Reed, Lowell J., 162, 168, 465, 476
 Regression, 103, 132
 and analysis of variance, 127, 157, 273, 348
 average within groups, 327
 coefficient, b , 106
 partial, $b_{Y1.2}$ or $b_{Y1.23}$, 340, 367
 standard deviation of, 119, 349, 368
 standard partial, $b'_{Y1.2}$, 342
 and correlation, 142, 155, 185, 372
 curvilinear, 374
 deviation from, 107, 113
 in each group, 327
 equation, 323
 linear, 107
 multiple, 340
 polynomial
 2nd degree, 379
 orthogonal, 388
 in experimentation, 126
 of group means, 328
 individual comparisons in, 409
 linear, 346
 multiple, 340
 correlation method, 340, 372
 direct method, 364, 372
 plane, 340
 in randomized blocks, 415
 and rates, 134
 values, 257
 Regression Y , $\hat{Y}$, 107
 Relative frequency, 431
 Relative growth rate, 376
 Relative standard deviation, C , 41, 80
 Relative variation, 47
 Replications, 43, 265, 268, 271, 275, 308
 Representative; *see* Sample
 Reversal experiment, 421
 Richardson, C. H., 116, 137, 199, 200, 213, 227, 252
 Rider, Paul R., 74
 Rietz, Henry Lewis, 185, 187
 Rounding, 94
 Running, T. R., 374, 378, 399

 S , sum or summation, 17
 S_n , sum of n th powers of deviations, 175
 $\sum X$, sum of observed values, 34
 s , standard deviation, 36
 s^2 , s_x^2 , $s_{y.2}$, etc.; *see* Variance
 s_b , standard deviation of regression coefficient, 119
 s_n , sum of n th powers, 175
 s_z , standard deviation of mean, 41, 61
 s_Y , standard error of Y , 120
 $s_{\hat{Y}}$, standard error of $\hat{Y}$, 120
 $s_{Y.12}$, standard error of estimate in multiple regression, 347
 $s_{Y.2}$, standard error of estimate in linear regression, 118
 Salmon, S. C., 46, 53
 Sample
 large, 169
 normally distributed, 55
 random, 2, 9, 42, 49, 57, 267, 455
 of means, 57
 from a normal distribution, 57
 representative, 1, 454
 small, 38, 175
 stratified, 2, 454
 systematic, 2, 267
 Sample size, n , 7, 24, 26, 46, 79, 456
 Sampling area, 468
 Sampling design, 453
 Sampling error, 466, 468
 Sampling from a normally distributed population, 52, 54, 57, 75
 Sampling unit, 467
 Sanders, H. G., 339
 Schott, R. G., 27, 30
 Season, 301
 Second degree polynomial, 379
 Second degree term, 379
 Semilogarithmic paper, 375
 Sets of comparisons
 nonorthogonal, 406
 orthogonal, 403
 in regression, coefficients for, 410
 Sex ratio, 27
 Sheppard, W. F., 173, 187
 Sheppard's correction, 173
 Shive, J. W., 93, 102
 Short cuts, 89
 in regression, 121
 individual comparisons, 400
 Sigma, σ , standard deviation, parameter, 39
 Significance, test of; *see* Hypothesis, test of
 Significance in covariance, 319, 327, 329
 Significance of
 chi-square, 22, 182
 correlation coefficient, 149
 departure from linearity, 381, 382, 384

Significance of — *Continued*

- difference between two
 - adjusted means, 330
 - means, 77, 226, 263
 - partial regression coefficients, 368
 - ratios, 195
 - regressions, 330
 - differences among
 - adjusted means, 319, 330
 - means, 221, 261
 - regression coefficients, 327
- F , 220
- mean difference, 44
- multiple correlation coefficient, R , and
 - multiple regression, 349
- partial correlation coefficient, 358
- partial regression coefficient, 349, 366
- R , 349
- range, 97, 409
- regression coefficient, 119
- standard partial regression coefficient, 349
- Significant digits, 95
- Significant difference between
 - means, 406
 - sums, 426
- Single degree of freedom, 400
- Size of sample; *see* Sample size
- Skewness; *see* Distribution, asymmetrical
- Slope of line, 107
- Smith, Arthur H., 82, 88, 385, 399
- Smith, B. Babington, 165, 168
- Smith, C. E., 200, 213
- Smith, F. B., 82, 88, 246, 252, 275, 316
- Smith, Stewart N., 79, 88
- Smoothing, 374
- Snedecor, George W., 111, 126, 137, 141, 168, 206, 213, 219, 235, 252, 262, 295, 317, 324, 338, 339, 350, 359, 373
- Snedecor, James, 24, 30
- Snell, M. G., 141, 168
- Solid geometry, 340
- Spearman, C., 165, 168
- Split plot experiment, 309
- Square, latin, 271
- Square, mean; *see* Mean square
- Square root transformation, 446
- Standard deviate, 144
- Standard deviation, s , 36, 38, 40
 - calculation
 - from deviations, 37
 - in large samples, 169
 - standard method, 89
 - with calculating machine, 91
 - characteristics, 38
 - of difference, s_D , 67, 76, 181
 - estimate of parameter, 60
 - from regression, $s_{y \cdot x}$, 117, 347
 - of g_1 and g_2 , 175, 177
 - of mean, $s_{\bar{x}}$, 41, 61
 - pooled estimate of, 76, 215
 - of regression coefficient, 119, 349, 368

Standard deviation — *Continued*

- relative; *see* Coefficient of variation
- of $\bar{Y}$, 120, 369
- Standard error; *see* Standard deviation
- Standard measure, 145, 181, 361
- Standard partial regression coefficient, $b'_{Y1 \cdot 2}$, 342
- Stark, Arvil L., 125, 137
- Statistic, 38, 54, 59, 60, 61, 62, 73, 76, 106, 117, 146, 181, 349, 433
- Statistical control, 126
- Statistical evidence; *see* Evidence
- Statistical tables, 19
- Stewart, R. T., 108, 137
- Strand, Norman V., 204, 213, 291, 295, 317
- Strata (stratification), 72, 454, 464
 - effectiveness of, 473
- Stuart, Neil W., 96, 102
- "Student," 38, 53, 267, 317, 441, 445, 452
- Subclass; *see* Groups
- Subclass numbers
 - equal, 75, 214, 275
 - unequal, 80, 232, 240
 - proportional, 281, 295
 - disproportionate, 284
- Subdivision; *see also* Strata
 - hierarchical, 207
- Subsample, 212, 232, 431
- Sukhatme, P. V., 82, 88
- Sum
 - of products, 105, 108
 - variance of, 161
 - weighted, 384
- Sum of squares, 37
 - correction for, 89
 - and correlation, 157
 - of means, 215, 217
 - in regression, 127, 328, 350, 381, 388
 - for testing significance of adjusted means, 319
- Surface, F. M., 48, 53
- Swanson, Pearl P., 82, 88, 385, 399
- Switch-back experiment, 421
- Symbolical summary of computations
 - in analysis of variance, 219, 231, 234, 256, 261
- Symmetry, 55, 64, 147, 174, 433
- $t = (\bar{x} - m)/s_{\bar{x}}$, 42, 62, 119, 149, 175, 180, 227, 304, 349, 368, 406, 435
 - distribution of, 62
 - cumulative percentage, 64
 - equal to χ , 435
 - table of, 65
- $t^2 = F$, 227, 264
- Tabulating machines, 122
- Talley, Paul J., 35, 53
- Tam, R. K., 49, 53
- Technique, experimental
 - test of, 209

- Tedin, O., 267, 317
 Terminal value, 392
 Test of hypothesis; *see* Hypothesis
 Test of normality, 174, 176
 Test of null hypothesis; *see* Hypothesis
 Test of significance; *see* Hypothesis and
 also Significance
 Thompson, William, 52, 53
 Thornton, H. G., 210, 213
 Time series, 103
 Tippett, L. H. C., 98, 102
 Titus, Harry W., 339
 Total correlation; *see* Correlation
 Transformation, 316, 445, 448
 Treatment, dummy, 214, 260, 313
 Trial, uniformity, 214
 Triangle, Pascal's, 433
 2×2 table, 194, 284
 $2 \times 2 \times 2$ table, 200

 Unbiased estimate; *see* Estimate
 Uniformity trial, 214
 Unit of measurement, 467
 Universe, 460
 Unweighted means, method of, 287, 299

 Van Slyke, Donald D., 52, 53
 Variance, s^2 , 37, 186
 analysis of, 214
 components of, 228, 260, 272, 278, 307
 and correlation, 185, 348
 of deviations from regression, $s_{y \cdot x^2}$, 117,
 348, 368
 of difference between means, 181
 of estimate, $s_{y \cdot x^2}$, 118, 368
 estimate of; *see* Estimate
 homogeneity of, 249
 of mean, $s_{\bar{y}}^2$, 61
 pooled, 76
 of random differences, s_D^2 , 67
 of sums, s_B^2 , 161
 Variance ratio, 218
 Variate or variable
 continuous, 31, 171
 dependent, 105
 discrete or discontinuous, 40, 440
 independent, 105
 Variation or variability, 1, 9, 15, 31, 40,
 58, 67, 68, 70
 coefficient of; *see* Coefficient
 see also Standard deviation and Vari-
 ance

 Vasey, A. J., 266, 316
 Vos, B. J., 382

W or *w*, weight, 285, 463
 Wakeley, Ray E., 7, 30, 198, 213
 Walker, Helen M., 19, 30
 Walker, R. H., 78, 88, 210, 213
 Wallace, H. A., 126, 137, 350, 359, 373
 Warren, Ernest, 133, 137
 Watson, D. J., 455, 476
 Weighted average, 15
 Weighted squares of means, method of,
 291, 299
 Weighted sum, 384
 Weight, *w*, *W* or *N*, 106, 113, 285, 462,
 463, 464
 Welch, B. L., 339
 Wentz, J. B., 108, 137
 Wiebe, Gustav A., 54, 74, 171, 187
 Wilks, S. S., 74, 339
 Willcox, O. W., 378, 399
 Willier, J. G., 358, 373
 Wilsie, C. P., 311, 317
 Winsor, C. P., 304, 317, 451, 452
 Wishart, J., 268, 316, 339

X
 independent variate, 105
 observed value of variate or item, 16, 34
 x , deviation of *X* from mean, 34
 x' , standard deviate or measure, 144, 343
 $\bar{x}$, mean value of *X*, 34

Y, dependent variable, 105
 $\hat{Y}$, regression *Y*, 106
 y , deviation of $\hat{Y}$ from mean, 109
 $\hat{y}$, deviation of $\hat{Y}$ from $\bar{y}$, 110
 $\bar{y}$, mean of *Y*, 109
 $\hat{Y} - \bar{Y}$, deviation from regression or error
 of estimate, 110
 Yates, F., 23, 30, 194, 213, 219, 252, 264,
 268, 272, 285, 287, 301, 305, 317,
 354, 373, 410, 423, 430, 455, 466, 476
 Yates' correction or adjustment for con-
 tinuity, 22, 193
 Youden, W. J., 44, 53
 Yule, G. Udny, 16, 30, 164, 168, 214,
 252, 454, 476

z distribution, 150
 graph of, 152
 Zacopanay, I., 466, 476
 Zero order correlation; *see* Correlation

